

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA E GESTÃO DO CONHECIMENTO**

Thales do Nascimento da Silva

**UM MODELO BASEADO EM ONTOLOGIA PARA SUPORTE A
TAREFA INTENSIVA EM CONHECIMENTO DE
RECOMENDAÇÃO**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina para a obtenção do Grau de Mestre em Engenharia e Gestão do Conhecimento.
Orientador: Prof. Dr. Alexandre Leopoldo Gonçalves.
Coorientador: Prof. Dr. João Artur de Souza.

Florianópolis
2015

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária
da UFSC

Silva, Thales do Nascimento da

Um Modelo Baseado em Ontologia para Suporte a Tarefa
Intensiva em Conhecimento de Recomendação / Thales do
Nascimento da Silva; orientador, Alexandre Leopoldo
Gonçalves; coorientador, João Artur de Souza. -
Florianópolis, SC, 2015.

149 p.

Dissertação (mestrado) - Universidade Federal de Santa
Catarina, Centro Tecnológico. Programa de Pós-Graduação em
Engenharia e Gestão do Conhecimento.

Inclui referências

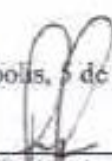
1. Engenharia e Gestão do Conhecimento. 2. Sistemas de
Recomendação. 3. Representação do Conhecimento. 4.
Ontologia. I. Gonçalves, Alexandre Leopoldo . II. Souza,
João Artur . III. Universidade Federal de Santa Catarina.
Programa de Pós-Graduação em Engenharia e Gestão do
Conhecimento. IV. Título.

Thales do Nascimento da Silva

**UM MODELO BASEADO EM ONTOLOGIA PARA SUPORTE A
TAREFA INTENSIVA EM CONHECIMENTO DE
RECOMENDAÇÃO**


Esta Dissertação foi julgada adequada para obtenção do Título de "Mestre em Engenharia e Gestão do Conhecimento", e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento.

Florianópolis, 5 de março de 2015.

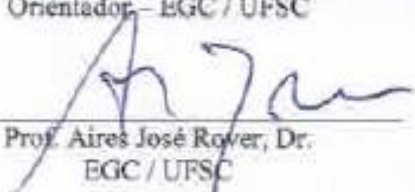


Prof. Roberto Carlos dos Santos Pacheco, Dr.
Coordenador do Curso


Banca Examinadora:




Prof. Alexandre Leopoldo Gonçalves, Dr.
Orientador – EGC / UFSC



Prof. Aires José Rover, Dr.
EGC / UFSC



Prof. Aran Bey Tcholakian-Morales, Dr.
EGC / UFSC



Prof. Eros Comunello, Dr.
MCA / Univali

Dedico este trabalho aos meus queridos pais, José e Denise e a minha companheira Pâmela. Obrigado pela paciência e carinho em todos os momentos desta caminhada.

AGRADECIMENTOS

Muitas pessoas contribuíram para o desenvolvimento deste trabalho. Quero neste momento apresentar meus sinceros agradecimentos. Agradeço a Deus por me dar força e inspiração. A Pâmela por compreender o tempo ausente, pelo carinho e dedicação. Aos meus pais que sempre me incentivaram, e como professores me ensinaram o valor da educação. A meu orientador Alexandre Leopoldo Gonçalves que foi fundamental para a conclusão desta dissertação através de sua imensa disponibilidade e sugestões que nortearam a pesquisa. Gostaria de agradecer também a meu coorientador, João Artur de Souza, e os demais professores do EGC por compartilhar suas experiências e conhecimentos.

O conhecimento serve para encantar as pessoas, não
para humilhá-las.

Mario Sergio Cortella

RESUMO

A evolução da Web tem propiciado uma mudança no comportamento de seus usuários, de consumidores para produtores de informação. Esta mudança trouxe consigo alguns desafios, dentre eles, o aumento exponencial de conteúdo, o que em muitos contextos promove dificuldades de escolha por parte dos usuários. A partir deste cenário iniciaram-se estudos com a finalidade de contribuir na filtragem do excesso de informação, surgindo os Sistemas de Recomendação. Com a evolução dos Sistemas de Recomendações surgiram diversas abordagens com o intuito de atender as demandas existentes. Uma dessas abordagens utiliza-se de estruturas semânticas sendo classificada como Sistemas de Recomendação baseados em Conhecimento. Em sua maioria fazem uso de ontologias e realizam recomendações a partir do conhecimento de um domínio específico. Porém, utilizadas de forma isolada as ontologias de domínio não possuem a capacidade de suportar o processo de recomendação como um todo. Diante disso, o presente trabalho propõe um modelo de representação do conhecimento capaz de lidar com a tarefa intensiva em conhecimento de recomendação. A partir desta premissa uma metodologia de construção de ontologias foi utilizada para modelar uma ontologia de domínio. Adicionalmente, foi proposta uma ontologia de tarefa de modo que tarefas genéricas possam ser especificadas. A integração das ontologias possibilitou que algumas abordagens de recomendação fossem executadas. O modelo proposto demonstra-se consistente uma vez que foi capaz de responder as questões de competências elaboradas na fase de modelagem. Apresenta também, a capacidade de tratar itens de forma genérica, podendo ser aplicado em diversos contextos. Ressalta-se ainda que a integração entre as ontologias de domínio e tarefa se mostra adequada, propiciando recomendações que visam auxiliar no processo de decisão.

Palavras-chave: Sistemas de Recomendação. Representação do Conhecimento. Ontologia.

ABSTRACT

The evolution of the Web has promoted a change in the behavior of its users from consumers to producers of information. This change brought with it some challenges, among them, the exponential growth of content which in many contexts generates difficulties of choice for users. From this scenario studies were started in order to contribute on information filtering, so emerging the Recommender Systems. With the development of it appeared several approaches to meet the existing demand. One of these approaches makes use of semantic structures being classified as Knowledge-based Recommender Systems. Mostly makes use of ontologies and carry out recommendations from the knowledge of a specific domain. However, if used alone domain ontology does not have the ability to support the recommendation process as a whole. Therefore, this work proposes a knowledge representation model able to handle the knowledge intensive task on recommendation. From this premise a methodology for building ontology was used to build a domain ontology. In addition, we propose an ontology task aiming to specify tasks in a generic way. The ontologies integration enabled that some recommendation approaches were implemented. The proposed model has demonstrated consistency since it was able to answer the competence questions elaborated in the modeling phase. It also shows the ability to process items generally, supporting its application in different contexts. Finally, we state that the integration of domain and task ontologies seems adequate providing recommendations to assist in decision making process.

Keywords: Recommender Systems. Knowledge Representation. Ontology.

LISTA DE FIGURAS

Figura 1- Sequência de etapas para a realização da dissertação.....	33
Figura 2 - Funcionamento genérico de um sistema de recomendação.	37
Figura 3 - Processo de recomendação baseada em conteúdo.	42
Figura 4 - Similaridade, diferenças e oportunidades.	45
Figura 5 - Tipos de ontologias.	61
Figura 6 - Fragmento da ontologia de reparação de danos causados por petróleo.	65
Figura 7 - Determinando o tamanho da média (visão da operação).	65
Figura 8 - Exemplo de árvore de decomposição de tarefa.	66
Figura 8 - Esquema resumindo com os passos da Metodologia 101.....	67
Figura 9 - Estrutura hierárquica das classes do modelo.	75
Figura 10 - Fragmento da classe Interacao	78
Figura 11 - Fragmento da classe Categoria	79
Figura 12 - Fragmento da classe Grupo	80
Figura 13 - Fragmento da classe Usuario	81
Figura 14 - Fragmento da classe Item	82
Figura 15 - Estrutura da ontologia proposta.....	83
Figura 16 - Instâncias da classe Categoria	88
Figura 17 - Instâncias da classe Item	89
Figura 18 - Instâncias da classe Grupo	90
Figura 19 - Instâncias da classe Usuario	91
Figura 20 - Instâncias da classe Numerica	91
Figura 22 - Estrutura hierárquica das classes da ontologia de tarefa.	92
Figura 23 - Estrutura da ontologia de tarefa.....	95
Figura 24 - Instâncias de uma tarefa.	95
Figura 25 - Descrição detalhada do modelo.....	97
Figura 26 - Site de recomendações de filmes MovieLens®.	98
Figura 27 - Site de recomendações de filmes MovieLens (Filme <i>The Godfather</i>).	99
Figura 28 - Estrutura da base MovieLens 100k.	101
Figura 29 - Instâncias da classe Categoria	106
Figura 30 - Exemplo de instância da classe Item	107
Figura 31 - Instâncias da classe Grupo	107
Figura 32 - Relacionamentos entre usuários e grupos.....	108
Figura 33 - Relacionamentos entre usuários através da similaridade.....	108
Figura 34 - Fluxo, interações e componentes envolvidos para o processamento de recomendações.....	116
Figura 35 - Estrutura genérica para execução de tarefas.....	117
Figura 36 - Estrutura da tarefa “ <i>Recomendar Itens</i> ” (Filtragem Colaborativa)..	118
Figura 37 - Usuários mais similares.....	120
Figura 38 - Usuários menos similares.....	121
Figura 39 - Usuários menos similares ao USUARIO_1.	121

Figura 40 - Usuários mais similares ao USUARIO_1.	121
Figura 41 - Os cinco usuários mais similares ao USUARIO_1.	122
Figura 42 - Lista de itens com seus valores de predição.	122
Figura 43 - Visão geral das recomendações na ontologia de domínio.	123
Figura 44 - Estrutura da tarefa “Recomendar Itens” (Demográfica).	124
Figura 45 - Estrutura da tarefa “Recomendar para Mulheres”.	126

LISTA DE TABELAS

Tabela 1 - Relação de interações entre usuários e itens.	44
Tabela 2 - Usuários, itens e suas avaliações.	52
Tabela 3 - Similaridades dos usuários obtidas através da distância euclidiana.	52
Tabela 4 - Similaridades normalizadas dos usuários obtidas através da distância euclidiana.	53
Tabela 5 - Similaridades dos usuários obtidas através da correlação de Pearson.	54
Tabela 6 - Similaridades dos usuários obtidas através do cosseno.	55
Tabela 7 - Bases de dados disponibilizadas pelo GroupLens.	100
Tabela 8 - Estrutura de um registro (filme) contido na tabela <i>ITEMS</i>	101
Tabela 9 - Estrutura de um registro (usuário) contido na tabela <i>USERS</i>	103
Tabela 10 - Estrutura de um registro (avaliação) contido na tabela <i>RATINGS</i>	104
Tabela 11 - Comparação entre os conjuntos de dados, completo e reduzido.	105
Tabela 12- Resultado da consulta do Quadro 22.	110
Tabela 13 - Resultado da consulta do Quadro 23.	110
Tabela 14 - Resultado da consulta do Quadro 24.	111
Tabela 15 - Resultado da consulta do Quadro 25.	112
Tabela 16 - Resultado da consulta do Quadro 26.	112
Tabela 17 - Resultado da consulta do Quadro 27.	113
Tabela 18 - Resultado da consulta do Quadro 28.	114
Tabela 19 - Resultado da consulta do Quadro 29.	115
Tabela 20 - Os 10 itens mais bem avaliados (Geral).	127
Tabela 21 - Os 10 itens mais bem avaliados (Mulheres)	127
Tabela 22 - Os 10 itens mais bem avaliados (Homens).	127
Tabela 23 - Os 10 itens mais bem avaliados (Menores de 25 anos).	128
Tabela 24 - Os 10 itens mais bem avaliados (De 25 a 50 anos).	128
Tabela 25 - Os 10 itens mais bem avaliados (Maiores de 50 anos).	129
Tabela 26 - Os 10 itens mais bem avaliados (Mulheres com mais de 50 anos).	129

LISTA DE QUADROS

Quadro 1 - Abordagens de Sistemas de Recomendação.	41
Quadro 2 - Visão geral da abordagem demográfica.	46
Quadro 3 - Visão geral da abordagem baseada em conhecimento.	49
Quadro 4 - Visão geral da abordagem baseada em utilidade.	49
Quadro 5 - Proposta de classificação de Ontologias.	61
Quadro 6 - Repositórios de ontologias utilizados na pesquisa.	72
Quadro 7 - Lista inicial de termos relevantes.....	73
Quadro 8 - Lista de classes obtidas.....	74
Quadro 9 - Alguns tipos primitivos de dados.	76
Quadro 10 - Propriedade de dado do modelo.	77
Quadro 11 - Exemplos de cardinalidade mínima, exata e máxima.	84
Quadro 12 - Cardinalidade das propriedades do modelo.	85
Quadro 13 - Restrições de propriedades da classe Usuario	86
Quadro 14 - Restrições de propriedades da classe Item	86
Quadro 15 - Restrições de propriedades da classe Categoria	87
Quadro 16 - Restrição de propriedade da classe Interacao	87
Quadro 17 - Restrição de propriedade da classe Recomendacao	88
Quadro 18 - Lista de classes da ontologia de tarefa.	92
Quadro 19 - Lista de propriedades de dado das classes da ontologia de tarefa.	93
Quadro 20 - Adequação do conjunto de dados para o modelo proposto (data de nascimento).	104
Quadro 21 - Adequação do conjunto de dados para o modelo proposto (Categoria preferida).	105
Quadro 22 - Consulta em linguagem SPARQL relativa à pergunta 5.4.1.	110
Quadro 23 - Consulta em linguagem SPARQL relativa à pergunta 5.4.2.	110
Quadro 24 - Consulta em linguagem SPARQL relativa à pergunta 5.4.3.	111
Quadro 25 - Consulta em linguagem SPARQL relativa à pergunta 5.4.4.	111
Quadro 26 - Consulta em linguagem SPARQL relativa à pergunta 5.4.5.	112
Quadro 27- Consulta em linguagem SPARQL relativa à pergunta 5.4.6.	113
Quadro 28 - Consulta em linguagem SPARQL relativa à pergunta 5.4.7.	114
Quadro 29 - Consulta em linguagem SPARQL relativa à pergunta 5.4.10.	114
Quadro 30 - Detalhamento da tarefa apresentada na Figura 36.	119
Quadro 31 - Detalhamento da tarefa apresentada na Figura 44.	124

LISTA DE ABREVIATURAS E SIGLAS

IMDb - *Internet Movie Database*

OWL - *Web Ontology Language*

RDF - *Resource Description Framework*

SPARQL - *SPARQL Protocol and RDF Query Language*

SR - *Sistemas de Recomendação*

W3C - *World Wide Web Consortium*

SUMÁRIO

1 INTRODUÇÃO	25
1.1 PROBLEMÁTICA	28
1.2 OBJETIVOS	29
1.2.1 Objetivo Geral	29
1.2.2 Objetivos Específicos	29
1.3 JUSTIFICATIVA	30
1.4 ESCOPO DO TRABALHO	31
1.5 ADERÊNCIA AO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA E GESTÃO DO CONHECIMENTO	32
1.6 METODOLOGIA	32
1.7 ESTRUTURA DO TRABALHO	34
2 SISTEMAS DE RECOMENDAÇÃO	36
2.1 BENEFÍCIOS DOS SISTEMAS DE RECOMENDAÇÃO	38
2.2 ABORDAGENS DE RECOMENDAÇÃO	40
2.2.1 Baseada em Conteúdo	42
2.2.2 Filtragem Colaborativa	43
2.2.3 Demográfica	45
2.2.4 Baseada em Comunidade	46
2.2.5 Híbrida	47
2.2.6 Baseada em Conhecimento	48
2.2.7 Baseada em Utilidade	49
2.2.8 Baseada em Crítica	50
2.2.9 Outras Abordagens	50
2.3 MEDIDAS DE SIMILARIDADE	51
2.3.1 Distância Euclidiana	52
2.3.2 Correlação de Pearson	53
2.3.4 Cosseno	55
2.3.5 Coeficiente de Jaccard	56
2.4 PROBLEMAS DOS SISTEMAS DE RECOMENDAÇÃO	56
2.5 CONSIDERAÇÕES FINAIS	58
3 REPRESENTAÇÃO DO CONHECIMENTO	58
3.1 ONTOLOGIAS	59
3.2 ONTOLOGIAS DE TAREFA	64
3.2 METODOLOGIAS DE ENGENHARIA DE ONTOLOGIA	66
3.2.1 Metodologia 101	66

3.2.2 Metodologia METHONTOLOGY	68
3.2.3 Metodologia KACTUS	68
3.2.4 Metodologia TOVE	69
3.3 CONSIDERAÇÕES FINAIS	70
4 MODELO PROPOSTO	70
4.1 DETERMINANDO O ESCOPO DA ONTOLOGIA	70
4.2 CONSIDERAR O REUSO	72
4.3 LISTAR TERMOS IMPORTANTES	72
4.4 DEFINIR CLASSES	74
4.5 DEFINIR PROPRIEDADES	75
4.5.1 Fragmento da classe Interacao	78
4.5.2 Fragmento da classe Categoria	79
4.4.4 Fragmento da classe Grupo	79
4.4.5 Fragmento da classe Usuario	80
4.4.5 Fragmento da classe Item	82
4.4.1 Ontologia de domínio	82
4.6 DEFINIR RESTRIÇÕES DAS PROPRIEDADES	83
4.6.1 Propriedades de Dados	83
4.6.2 Propriedades de Objeto	85
4.7 CRIAR INSTÂNCIAS	88
4.8 ONTOLOGIA DE TAREFA	92
4.9 CONSIDERAÇÕES FINAIS	96
5 APLICAÇÃO DO MODELO E DISCUSSÃO DOS RESULTADOS	97
5.1 DESCRIÇÃO DO CENÁRIO DE AVALIAÇÃO	98
5.2 POPULANDO A ONTOLOGIA	104
5.3 VISUALIZANDO INSTÂNCIAS	106
5.4 RESPONDENDO AS QUESTÕES DE COMPETÊNCIA	109
5.5 GERANDO RECOMENDAÇÕES	115
5.5.1 Recomendando através da Filtragem Colaborativa	117
5.5.2 Recomendando através da abordagem demográfica	123
5.6 CONSIDERAÇÕES FINAIS	130
6 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS	130
6.1 CONSIDERAÇÕES FINAIS	130
6.2 TRABALHOS FUTUROS	132
REFERÊNCIAS	134

1 INTRODUÇÃO

Inicialmente a Web era movida por um pequeno número de empresas e anunciantes que produziam conteúdo para um grande número de receptores (DEITEL; DEITEL, 2008). Sob esta perspectiva o usuário era um espectador, desprovido da capacidade de expressão. Este cenário praticamente não permitia a interação entre os consumidores e produtores de conteúdo, característica atualmente simples e muito habitual que possibilitou um grande avanço comercial e principalmente social na Internet.

No cenário inicial, também chamado de Web 1.0, os receptores ou usuários possuíam uma série de limitações. Dentre estas limitações destacam-se a baixa capacidade de personalização, controle quase exclusivo do emissor em relação ao conteúdo e pouca interação entre os receptores e o emissor (SAAD, 2008).

A mudança do papel do usuário foi a principal e a mais importante contribuição para a concepção de um novo paradigma da Web. O usuário antes um mero espectador, passou a participar ativamente na geração de conteúdo (JESUS; CUNHA, 2012). Esta mudança, que a princípio pode parecer simplória, proporcionou uma gama muito maior de possibilidades. Segundo Schneider, Conforto, Santarosa (2013) “o perfil do usuário da Web foi alterado com a conquista de novos instrumentos para gerar conhecimento, criar e interagir em comunidades digitais”. Diante deste cenário surge uma Web colaborativa que deixa de ser de mão única e é movida pela participação coletiva, interação e partilha (ARGOLLO et al., 2010). Santarosa, Conforto, Basso (2013) contribuem com esta afirmação expondo que neste novo paradigma é crescente o desenvolvimento de sistemas computacionais que agregam recursos de compartilhamento e interconexão de tecnologias e saberes.

Apesar da mudança de paradigma ser gradual, em 2004 Tim O'Reilly cunhou o termo mais aceito na literatura que denomina tal evolução, a Web 2.0 (O'REILLY, 2005). O termo foi apresentado em uma palestra cujo objetivo era discutir a autonomia do usuário na web. Desde então o mesmo foi amplamente utilizado e conceituado.

Diante desta nova realidade a Web proporciona diversas oportunidades para a criação de conhecimento através principalmente da interação entre usuários. O conhecimento pode ser gerado não só por meios textuais, mas também através da análise comportamental, das preferências e concepções de determinado grupo (SEGARAN, 2008).

Se por um lado a Web 2.0 proporciona diversas oportunidades, este cenário também lança desafios, entre estes, como filtrar as informações que são realmente úteis aos consumidores, levando em consideração suas preferências, já que analisar toda essa informação manualmente é inviável.

Para Cazella, Nunes e Reategui (2010) uma maneira tradicional de auxiliar na escolha dentre diversos itens oferecidos são as recomendações passadas por outras pessoas, pessoalmente ou através de textos de recomendação, opiniões de críticos, jornais, entre outros. Porém, à medida que cresce o número de sugestões e de itens disponíveis a escolha embasada nas sugestões deixa de ser prática (SEGARAM, 2008).

Como alternativa a esta necessidade destacam-se os Sistemas de Recomendação, sendo atualmente os *sites* de comércio eletrônico os maiores utilizadores destes sistemas que visam sugerir itens que se enquadrem e atendam as necessidades de seus clientes, tendo como objetivo principal o aumento da lucratividade (SCHAFFER; KONSTAN; RIEDL, 2001; MONTANER; LÓPEZ; ROSA, 2003).

Os Sistemas de Recomendação surgiram nos meados dos anos 90. Desde sua concepção até o momento os Sistemas de Recomendação tornaram-se uma importante área de pesquisa (GOLDBERG et al., 1992; RESNICK et al., 1994, SHARDANAND; MAES, 1995). A intenção destes sistemas consiste em simular o processo de sugestão utilizada comumente pelos indivíduos. Apesar de trivial, em alguns contextos o processo de análise e escolha de um item pode se tornar frustrante para os usuários. Tal fato pode ser ilustrado com a popularização do *E-commerce* em que o consumidor passou de alguns milhares de opções de livros disponíveis em lojas físicas para milhões de livros disponíveis em lojas online (SCHAFFER; KONSTAN; RIEDL, 2001).

Além do *E-commerce* (CAZELLA, NUNES, REATEGUI, 2010; GEDIKLI, JANNACH, GE, 2014; MORID, SHAJARI, HASHEMI, 2014) e a Web 2.0 (BOTHOS, APOSTOLOU, MENTZAS 2012;), a inteligência coletiva (BOTHOS, APOSTOLOU, MENTZAS 2012; SEGARAM, 2008) e a grande quantidade de informação disponível são argumentos utilizados para justificar os estudos recentes sobre os Sistemas de Recomendação.

Diante do principal objetivo destes sistemas, Ricci, Rokach e Shapira (2011) apresentam as funções que os mesmos devem desempenhar, sendo algumas delas: aumentar o número de itens vendidos, aumento da diversidade de itens vendidos, aumentar a

satisfação do usuário, aumentar a fidelidade do usuário, entender melhor o que o usuário quer.

Visado melhorar a qualidade das recomendações os Sistemas de Recomendação vêm evoluindo naturalmente principalmente na última década através de pesquisas acadêmicas e organizacionais (ADOMAVICIUS; TUZHILIN, 2005). Essas evoluções almejam atender diversos contextos que, na grande maioria das vezes possui certa complexidade envolvida no processo de sugestão. A partir disto surgiram diversas abordagens de recomendação, entre estas, destacam-se as citadas por Ricci, Rokach e Shapira (2011):

- **Baseada em conteúdo:** faz sugestões a partir da análise das características de itens que o usuário gostou, sugerindo itens com características semelhantes;
- **Filtragem colaborativa:** recomenda a um usuário itens que usuários semelhantes tenham gostado, porém ele ainda não tenha visto;
- **Baseada em conhecimento:** se utilizam de estruturas capazes de representar o conhecimento de determinado domínio para auxiliar no processo de recomendação;
- **Demográfica:** recomenda com base no perfil demográfico do usuário, fazendo recomendações com através de perfis demográficos semelhantes;
- **Baseada em comunidade:** faz recomendações com base nas preferências de usuário que pertence a mesma comunidade;
- **Híbrida:** utiliza a combinação de abordagens visando utilizar os pontos fortes de cada uma delas.

Além dessas abordagens, pode-se citar ainda a baseada em utilidade e a baseada em crítica. A primeira realiza recomendações a partir da utilidade de cada item para o utilizador (usuário) em que esta é calculada através do perfil do usuário (HUANG, 2011). Por outro lado, na abordagem baseada em crítica o usuário realiza diversas críticas em um processo interativo nas características dos itens sugeridos, assim gera-se uma nova sugestão a cada ciclo. O processo se repete até encontrar um item adequado (CHEN; PU, 2012).

Dentre as técnicas apresentadas a baseada em conteúdo, filtragem colaborativa, baseada em conhecimento e híbrida são as mais citadas na literatura. Os sistemas baseados em conhecimento destacam-se pela capacidade de sugerir itens com base em inferências sobre as

necessidades e preferências do usuário (BURKE, 2002). Com esse pressuposto é fundamental que, tanto as características dos produtos quanto as dos usuários estejam organizados em uma estrutura formal capaz de representar o domínio e suportar as inferências que serão aplicadas.

Entre as estruturas capazes de representar o conhecimento de domínio encontram-se as Ontologias que suportam características e relações de itens e usuários (BURKE; RAMEZANI, 2011). As Ontologias possuem a capacidade de representar o conhecimento, definindo conceitos e representando as relações semânticas de determinado domínio (GUARINO, 1998). Segundo Staab e Studer (2009) as Ontologias estão sendo cada vez mais utilizadas em Sistemas de Recomendação com intuito de utilizar técnicas baseadas em conhecimento para complementar as abordagens clássicas baseadas em Aprendizado de Máquina e Estatística.

1.1 PROBLEMÁTICA

Web 2.0 refere-se à segunda geração da *World Wide Web*. A transição para a fase atual da Web foi caracterizada pela mudança de páginas estáticas para conteúdo dinâmico, gerado pelo usuário, e o crescimento do conceito de Mídia Social. A Web 2.0 também evidenciou uma maior interação por parte dos usuários e o aumento da confiança na Internet. Esta evolução gerou um cenário em que a utilização e a dependência da Web estão intrínsecas em nosso cotidiano (GRANDSON, 2014). A possibilidade de geração de conteúdo pelo usuário aliada a uma evolução dos dispositivos computacionais vem promovendo um aumento na geração de informação.

Em uma análise preliminar a grande quantidade de conteúdo é benéfica, pois disponibiliza diversas opções, resultando na seleção de um conteúdo mais adequado diante as necessidades do usuário. Porém, a sobrecarga de informação pode frustrar o usuário no processo de seleção (BOLLEN, 2010). A partir deste pressuposto e com intuito de contornar esta dificuldade, surgem os Sistemas de Recomendação. Tais sistemas fazem sugestões de itens que melhor se adequem a determinado perfil de usuário.

Na atualidade, os Sistemas de Recomendação tem usufruído das tecnologias de Web Semântica para superar de forma eficaz os desafios relacionados com o incrível crescimento da Web, mais precisamente tratando a sobrecarga de informações e fontes de dados heterogêneas (COLOMBO-MENDOZA, 2015). Estes sistemas vêm ganhando

maturidade e dispõem de algumas abordagens distintas. A abordagem que utiliza ontologias para realizar recomendações é denominada baseada em conhecimento (CHEN et al., 2012).

Os Sistemas de Recomendação baseados em conhecimento que fazem uso de uma descrição semântica do domínio vem se tornando cada vez mais utilizada diante da perceptível melhora na qualidade das recomendações (COLOMBO-MENDOZA, 2015).

Estes sistemas têm a capacidade de auxiliar no processo de tomada de decisão, fazendo sugestões mais relevantes e com mais benefícios para quem o utiliza (WALTER et al., 2012). Outro aspecto de fundamental importância e que também pode ser utilizado para tornar uma organização mais competitiva é a capacidade de fornecer conhecimento realmente útil sobre os usuários e seus comportamentos (SEGARAM, 2008; WALTER et al., 2012).

Diante do grande potencial dos Sistemas de Recomendação surge um desafio no processo de concepção dos mesmos, questão que se refere à Engenharia do Conhecimento, mais precisamente no aspecto da modelagem do conhecimento. A construção de ontologias que suportem os processos de recomendação e suas tarefas tem se apresentado como um desafio para a área (WALK, 2012). Ricci, Rokach e Shapira (2011) também afirmam que a capacidade de moldar um Sistema de Recomendação genérico é relevante e adequado, pois a recomendação de itens de forma independentemente do domínio em que este esteja inserido é visto como um desafio (RICCI; ROKACH; SHAPIRA, 2011).

A partir destas afirmações apresenta-se a pergunta de pesquisa deste trabalho: **Como representar semanticamente o domínio de Sistemas de Recomendação de modo que seja possível suportar a tarefa intensiva em conhecimento de recomendação?**

1.2 OBJETIVOS

1.2.1 Objetivo Geral

Este trabalho possui como objetivo geral a proposição de um modelo de representação do conhecimento voltado à tarefa intensiva em conhecimento de recomendação.

1.2.2 Objetivos Específicos

Visando atingir o objetivo geral, alguns objetivos específicos são necessários, entre eles:

- Identificar os principais elementos dos Sistemas de Recomendação de modo a subsidiar a tarefa intensiva em conhecimento de recomendação;
- Propor um modelo baseado em Ontologia para representar o domínio dos Sistemas de Recomendação;
- Propor uma ontologia de tarefa visando promover suporte as diferentes abordagens atualmente disponíveis;
- Demonstrar a viabilidade do modelo proposto através de um cenário que possa ser aplicado aos Sistemas de Recomendação;
- Promover uma discussão dos resultados obtidos, bem como apresentar possibilidades de trabalhos futuros.

1.3 JUSTIFICATIVA

Os Sistemas de Recomendação estão em evidencia na atualidade. Estes sistemas estão em constante evolução e são alvo de diversos estudos e investimentos na atualidade. Um exemplo de grande repercussão na área foi o concurso promovido em 2006 pela Netflix®, empresa de locação virtual de vídeos, que ofereceu 1 milhão de dólares para quem conseguisse melhorar em 10% a estratégia que era utilizada pela empresa. O concurso chegou ao final em 2009 (LOHR, 2009).

O aumento da quantidade de itens, usuários, características e suas relações tem impacto direto na complexidade da recomendação (CAZELLA, NUNES, REATEGUI; 2010). A partir desta colocação pode-se afirmar que a recomendação é uma tarefa intensiva em conhecimento, que deve considerar diversos elementos e suas relações durante o processo de tomada de decisão.

Uma maneira de auxiliar esta tarefa é a utilização de Sistemas de Recomendação baseados em conhecimento. Estes sistemas utilizam estruturas semânticas, tais como as ontologias tornando possível, através de algoritmos de inferência, identificar a correlação entre as preferências de usuários e produtos existentes, serviços ou conteúdos (CARRER-NETO et al., 2012).

Porém, a utilização de ontologias na tarefa de recomendação promove alguns desafios, uma vez que estas possuem algumas limitações diante das necessidades da tarefa. Portanto, torna-se necessário aprimorar as ontologias para que realizem recomendações de forma adequada (POUCHARD; IVEZIC; SCHLENOFF, 2000; WALK,

2012). A capacidade de modelar ontologias de modo que possam dar suporte a execução das tarefas intensivas em conhecimento é um desafio importante e atual.

A construção de ontologias com foco na execução de tarefas é um tema pouco abordado na literatura, onde em grande parte das vezes os estudos apenas apresentam proposições de estruturas, mas não as colocam em prática. Esta abordagem cria uma lacuna entre a proposição e a implementação e demonstração prática de uma ontologia de tarefa.

Diante deste cenário o alinhamento e a adaptação da tarefa de recomendação com a utilização de ontologias evidenciam um grande desafio da área (WALK, 2012). Sendo assim, a proposta de um modelo que represente o conhecimento do domínio e suporte as tarefas exigidas no processo de recomendação oferece uma contribuição direta e atual no contexto dos Sistemas de Recomendação e da Representação de Conhecimento. A partir destes argumentos evidencia-se a justificativa e relevância da presente dissertação com contribuições no âmbito organizacional e acadêmico.

1.4 ESCOPO DO TRABALHO

Esta dissertação propõe um modelo baseado em Ontologia para a representação do domínio de recomendação com foco nas seguintes abordagens:

- Baseada em Conteúdo;
- Filtragem Colaborativa;
- Baseada em Conhecimento;
- Demográfica;
- Baseada em Comunidade;
- Híbrida;
- Baseada em Utilidade;
- Baseada em Crítica.

Além disso, o modelo proposto também inclui uma ontologia de tarefa composta por alguns métodos de resolução de problema que visam oferecer suporte as abordagens Demográfica, Baseada em Conteúdo, Baseada em Comunidade, Filtragem Colaborativa e Híbrida.

Este estudo não se estende ao desenvolvimento de *software* enquanto produto final e sim como um protótipo visando auxiliar na avaliação dos resultados obtidos. Não pertencem ao escopo a análise de

desempenho e a precisão das recomendações das técnicas selecionadas visto que existem diversos estudos na literatura que abrangem estas questões. Ressalta-se ainda, que este trabalho não possui foco na proposição de novas abordagens e técnicas no domínio de Sistemas de Recomendação.

1.5 ADERÊNCIA AO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA E GESTÃO DO CONHECIMENTO

O Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento – PPGEKC é composto por três áreas de concentração: Engenharia do Conhecimento, Gestão do Conhecimento e Mídia e Conhecimento. Esta dissertação pertence à área de Engenharia do Conhecimento.

A área em questão possui algumas linhas de pesquisa, dentre elas, a de Teoria e Prática da Engenharia do Conhecimento é a que melhor representa o trabalho realizado. Segundo EGC (2014) trabalhos desta linha “estudam-se metodologias e técnicas da Engenharia do Conhecimento e da Inteligência Computacional e suas relações com a gestão e com a mídia e conhecimento”.

Sendo assim a proposta envolve primordialmente a área de Engenharia do Conhecimento como subsídio a área de Gestão do Conhecimento. Isso se torna evidente visto que tem por objetivo a obtenção de melhorias nos resultados na tarefa de recomendação e representação do conhecimento envolvido. Por sua vez, tais melhorias visam a agregação de valor nas tarefas de criação, manutenção e disseminação do conhecimento por parte da Gestão do Conhecimento.

1.6 METODOLOGIA

Esta seção objetiva descrever a metodologia utilizada e classificar a pesquisa em diversos aspectos.

Este trabalho tem como natureza a pesquisa aplicada. Segundo Silva e Menezes (2001) a pesquisa de natureza aplicada “objetiva gerar conhecimentos para aplicação prática e dirigidos à solução de problemas específicos. Envolve verdades e interesses locais”.

Quanto a abordagem utilizada se caracteriza por uma pesquisa exploratória, que segundo Gil (1999) “têm como objetivo proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a constituir hipóteses”.

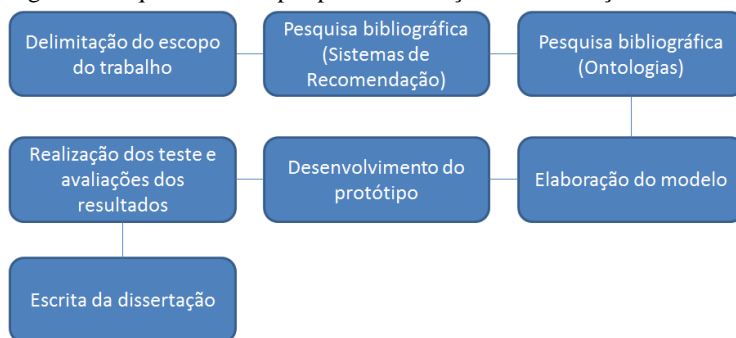
A pesquisa bibliográfica foi escolhida como procedimento técnico para o desenvolvimento desta dissertação. De acordo com Gil (1999) “A principal vantagem da pesquisa bibliográfica reside no fato de permitir ao investigador a cobertura de uma gama de fenômenos muito mais ampla do que aquela que poderia pesquisar diretamente”.

Para alcançar os objetivos desta pesquisa, o trabalho foi realizado através da execução dos seguintes passos:

- Realização das pesquisas bibliográficas que abordam Sistemas de Recomendação e Ontologias. Esta etapa subsidiou a construção do modelo e guiou os passos subsequentes;
- Proposição de um modelo capaz de atender aos objetivos deste trabalho;
- Desenvolvimento de um protótipo viabilizando a realização dos testes e a avaliação do modelo;
- Análise dos resultados obtidos;
- Apresentação das conclusões e perspectivas de trabalhos futuros.

A Figura 1 apresenta o procedimento utilizado para realização da presente pesquisa.

Figura 1- Sequência de etapas para a realização da dissertação.



Fonte: Autor.

É pertinente fazer um detalhamento dos procedimentos realizados, sendo assim:

- O modelo proposto foi modelado utilizando uma metodologia de construção de ontologia. Para

desenvolvimento do modelo foi utilizado o Protégé®, editor de ontologias, assim como, para detalhar graficamente a ontologia.

- Para a população da ontologia utilizou-se uma base de dados do domínio de recomendações de filmes, que contém informações sobre seus usuários, interações e itens.
- Os testes foram realizados a partir do modelo proposto contendo o conjunto de dados apresentado no item anterior. Primeiramente, analisou-se o comportamento da ontologia com os dados já inseridos. Avaliou-se também a capacidade da ontologia em responder as perguntas de competência. Por fim, a capacidade de gerar recomendações foi testada e avaliada.

1.7 ESTRUTURA DO TRABALHO

O presente trabalho é composto por seis capítulos descritos a seguir:

- **Capítulo 1:** capítulo introdutório que apresenta a problemática, os objetivos, justificativa, escopo do trabalho, aderência ao programa, metodologia e por fim a estrutura do trabalho.
- **Capítulo 2:** faz uma revisão bibliográfica sobre Sistemas de Recomendação. O capítulo inicia com uma breve introdução, e aborda os principais benefícios, algumas abordagens de recomendação, medidas de similaridade e os principais problemas da área.
- **Capítulo 3:** este capítulo abrange a Representação do Conhecimento, e promove uma visão geral sobre as Ontologias abordando também as ontologias de tarefas.
- **Capítulo 4:** objetiva descrever o modelo proposto através do detalhamento de todas as etapas envolvendo uma metodologia de construção de ontologia e desenvolvimento de uma ontologia de tarefa.
- **Capítulo 5:** realiza um detalhamento e aplicação do modelo proposto. Para isso descreve o cenário utilizado na avaliação, a população da ontologia, responde as perguntas de competência e por fim, gera algumas recomendações.
- **Capítulo 6:** por fim as considerações finais são apresentadas bem como as sugestões de trabalhos futuros.

2 SISTEMAS DE RECOMENDAÇÃO

Os Sistemas de Recomendação têm como objetivo realizar sugestões principalmente onde há um grande volume de opções, uma vez que nesta situação o processo de seleção pode se tornar difícil para o usuário (SENECAL; NANTEL, 2004).

Os Sistemas de Recomendação surgiram em meados dos anos 90 e desde então tem se tornado uma importante área de pesquisa (GOLDBERG et al., 1992; RESNICK et al., 1994, SHARDANAND; MAES, 1995). Segundo Adomavicius e Tuzhilin (2005) o interesse nesta área é elevado, pois possui relevantes problemas de pesquisas e muitas aplicações existentes.

O que torna este tipo de sistema útil, tanto para o usuário quanto para quem oferece o serviço é a capacidade de auxiliar a seleção de itens, tornando esta tarefa mais agradável e possivelmente promovendo melhores resultados. Diante destes argumentos Melville e Sindhvani (2010) afirmam que “O objetivo dos Sistemas de Recomendação (SR) é gerar recomendações válidas de itens que possam ser de interesse de um conjunto de usuários”, tendo em vista que para alcançar este objetivo existe uma série de abordagens e métodos capazes de fornecer recomendações personalizadas para os usuários (METTOURIS; PAPADOPOULOS, 2014). Em suma, a partir das características apresentadas dos Sistemas de Recomendação conclui-se que eles são ferramentas importantes no processo natural de tomada de decisão (MELVILLE; SINDHWANI, 2010).

A Figura 2 ilustra o funcionamento de um Sistema de Recomendação de maneira genérica, a fim de entender o funcionamento básico, sendo que nas próximas seções será realizada uma análise mais detalhada das abordagens e técnicas existentes.

Como é possível observar na Figura 2 o usuário é parte fundamental e participa ativamente no processo de recomendação, seja atuando apenas como visitante, ou como em alguns casos, fornecendo insumos para o funcionamento do sistema. Abordagens de recomendação como a filtragem colaborativa e demográfica utilizam diretamente o comportamento e as características dos usuários em seus processos de recomendação. Esse foco em realmente avaliar as particularidades de determinado usuário ou grupo é uma tendência nos sistemas de recomendações atuais, onde o objetivo é gerar recomendações altamente personalizadas fazendo com que cada usuário receba recomendações distintas e provavelmente mais adequadas (RICCI; ROKACH; SHAPIRA, 2011; TORRES, 2004).

Figura 2 - Funcionamento genérico de um sistema de recomendação.



Fonte: Autor.

Outro ponto importante nos Sistemas de Recomendação é o conceito de item. Em linhas gerais um item é o conteúdo da recomendação que é proposta ao usuário. Segundo Ricci, Rokach e Shapira (2011) "Item" é o termo geral usado para designar o que o sistema recomenda aos usuários. O SR normalmente se concentra em um tipo específico de item (por exemplo, CDs, ou notícias) e, conseqüentemente, o projeto, a interface gráfica a disposição do usuário, e a abordagem de recomendação usada para gerar as recomendações são personalizadas de maneira que seja possível fornecer sugestões úteis e eficazes para um tipo específico de item.

O conceito de recomendação de itens é utilizado de maneira genérica, uma vez que uma recomendação pode ser aplicada as mais diversas áreas, podendo representar livros, carros, passagens aéreas, pessoas, filmes, entre outros. Cazella, Nunes e Reategui (2011) resumem afirmando que um item pode ser um produto, serviço ou pessoa. Tal afirmação adequa-se bem a realidade atual abrangendo não somente produtos e serviços, mas também pessoas, características fundamentais na Web atual, principalmente devido às redes sociais.

A recomendação de itens é realizada a partir de abordagens de recomendação, como pode ser observado na Figura 2. Na atualidade existem muitas abordagens ou técnicas que podem ser utilizadas, e seu

número cresce rapidamente. O desenvolvimento de técnicas distintas é impulsionado pelos diversos cenários em que estes sistemas podem estar inseridos, gerando assim um aperfeiçoamento para determinada realidade ou até mesmo o desenvolvimento de novas técnicas que tenham maior precisão perante as necessidades específicas.

2.1 BENEFÍCIOS DOS SISTEMAS DE RECOMENDAÇÃO

Os Sistemas de Recomendação podem promover diversos benefícios. Porém, existem duas visões distintas quanto aos possíveis benefícios: a de quem implementa ou oferece o serviço de recomendação e a do usuário que utiliza esse sistema.

Na maioria das vezes a visão do prestador de serviço é mais comercial, porém vale lembrar que o objetivo da recomendação nem sempre é vender um produto ou serviço, tendo como exemplo as redes sociais (por exemplo, recomendação de amigos) ou recomendações de funcionários para um novo projeto. Apesar desta ressalva na literatura a grande maioria dos autores tem uma visão mercadológica quando se referem aos Sistemas de Recomendação. Essa tendência pode ser observada na percepção de Schroder et al. (2011) quando afirmam que as razões mais comuns para a implementação de um sistema de recomendação é o desejo de aumentar a satisfação do usuário e consequentemente aumentar o sucesso econômico da plataforma. Essa visão mais comercial deve-se ao fato do setor de *e-commerce* ser o maior utilizador desse tipo de sistema.

Diante desta realidade é possível citar um estudo realizado por Ricci et al. (2011) que identifica e descreve as tarefas mais importantes executadas pelos Sistemas de Recomendação na atualidade. As tarefas apresentadas segundo a visão dos autores são:

- **Aumentar o número de itens vendidos:** os itens são direcionados para cada perfil de usuário, aumentando assim as vendas. No caso de sistemas não comerciais é possível aumentar o acesso a um site, fazer com que determinada notícia seja mais lida, entre outras possibilidades. Esta função é considerada a mais importante dos Sistemas de Recomendação;
- **Diversificar as vendas:** também é considerada uma importante função, pois sugere não só os itens mais populares, mas sim os mais relevantes de acordo com o

perfil do usuário. Isso faz com que uma quantidade mais ampla de itens sejam selecionados;

- **Aumentar a satisfação do usuário:** recomendações mais relevantes tornam a experiência do usuário mais agradável fazendo com que consequentemente ele utilize mais o sistema e se sinta satisfeito com suas escolhas;
- **Aumentar a fidelidade do usuário:** um usuário satisfeito tende a utilizar mais vezes o sistema. A partir desta fidelização e maior utilização é possível criar um perfil mais preciso e fazer recomendações cada vez mais relevantes;
- **Perceber as necessidades do usuário:** entender as preferencias do usuário pode ser extremamente útil, podendo auxiliar na gestão organizacional. Os usuários podem fornecer informação de forma explicita através de questionários ou de forma implícita através das características comportamentais expostas pelo usuário durante a utilização do sistema.

Ainda que o estudo de Ricci et al. (2011) possua um viés mercadológico, as tarefas apresentadas não limitam-se a vender um item, sendo que tais conceitos podem ser facilmente generalizados para domínios não comerciais, como por exemplo, a recomendação de artigos científicos a partir do perfil de interesse de determinado usuário.

Além das tarefas já apresentadas, Herlocker et al. (2004) apresentam uma lista de tarefas que podem ser desempenhadas pelos sistemas de recomendação, numa visão mais voltada à percepção dos usuários, entre elas:

- **Encontrar alguns itens adequados:** recomendar ao usuário alguns itens ordenados por relevância;
- **Encontrar todos os itens adequados:** recomendar ao utilizador todos os itens que satisfazem as suas necessidades;
- **Destacar itens num contexto:** considerando um determinado contexto o sistema deve destacar uma lista de itens baseando-se no histórico de preferências do utilizador;
- **Encontrar uma sequência adequada:** recomendar uma sequência de itens que no todo interessam ao utilizador;
- **Encontrar um pacote de itens adequado:** encontrar um conjunto de itens que agregados interessam ao utilizador;

- **Auxiliar a navegação:** ajudar o utilizador a navegar pelos diversos itens que de algum modo possam satisfazer as suas intenções naquela sessão;
- **Criar confiança:** disponibilizar ao utilizador mais cético um conjunto de funcionalidades que o ajudem a testar a qualidade das recomendações;
- **Melhorar o perfil:** ajudar e incentivar o utilizador a revelar os seus gostos e preferências;
- **Auxiliar a expressão:** proporcionar satisfação ao utilizador que se sente realizado por expressar o seu conhecimento;
- **Ajudar terceiros:** proporcionar satisfação ao utilizador que se sente realizado por contribuir para o bem comum, sabendo que à partida estará a ajudar terceiros e não a si mesmo;
- **Influenciar terceiros:** existem utilizadores de sistemas de recomendação que têm como única função influenciar outros a adquirirem determinados itens. Esta ação pode não ser vista como bem intencionada, mas está contemplada neste tipo de sistema.

Apesar de os Sistemas de Recomendação terem sido desenvolvidos como uma solução para o problema da grande quantidade de opções, sua evolução vem beneficiando, tanto organizações, quanto usuários (BURKE; FELFERNIG; GÖKER, 2011).

Vale salientar que um sistema de recomendação pode e, na maioria das vezes, não oferece todos os benefícios apresentados por Herlocker et al. (2004). Esse fato se deve a diversidade e foco que possuem. Sendo assim, muitas vezes alguns benefícios podem não serem relevantes em determinado contexto.

Diante da exposição dos principais benefícios, a seguir são apresentadas as principais abordagens utilizadas pelos Sistemas de Recomendação.

2.2 ABORDAGENS DE RECOMENDAÇÃO

Atualmente a literatura apresenta diversas abordagens de recomendação. Trabalhos como o de Friedrich e Zanker (2011) utilizam o termo “paradigma” na categorização dos Sistemas de Recomendação enquanto outros (CAZELLA; NUNES; REATEGUI, 2010; BURKKE, 2002) consideram as diversas propostas de recomendação como

“abordagem”. A gama de abordagens atuais contém muitas especializações e derivações desenvolvidas para melhor atender uma demanda específica. Diante deste pressuposto justifica-se a utilização do termo “abordagens de recomendação” no presente trabalho.

O número e os tipos de abordagens existentes não é consensual. Na Quadro 1, Burke (2002) apresenta um conjunto composto por cinco abordagens amplamente referenciadas, onde **I** é o conjunto de itens passíveis de recomendação, **U** é o conjunto de usuários com preferências conhecidas, **u** é o usuário alvo da recomendação, e **i** é um item que, se considerado adequado, pode ser recomendado para **u**.

Quadro 1 - Abordagens de Sistemas de Recomendação.

Abordagem	Base	Entrada	Processo
Filtragem Colaborativa	Avaliações de U para itens em I .	Avaliações de u para itens em I	Identificar usuários do conjunto U similares a u , e verificar a diferença dos conjuntos.
Baseada em conteúdo	Características de itens em I .	Avaliações de u para itens em I	Classificar o comportamento de avaliação de u e utilizar em i .
Demográfica	Informações demográficas sobre U e avaliações desses usuários de itens em I .	Informações demográficas sobre u	Identificar usuários demograficamente similares a u , e verificar a diferença dos conjuntos.
Baseada em utilidade	Características de itens em I .	A função de utilidade de itens em I e a descrição de preferências de u .	Aplicar uma função e gerar um classificação de determinado i .
Baseada em conhecimento	Características de itens I . Conhecimento de como esses itens se adequam as necessidades do usuário	A descrição das necessidades ou interesses de u .	Inferir sobre a correspondência entre i e as necessidades de u .

Fonte: Adaptado de Burke (2002).

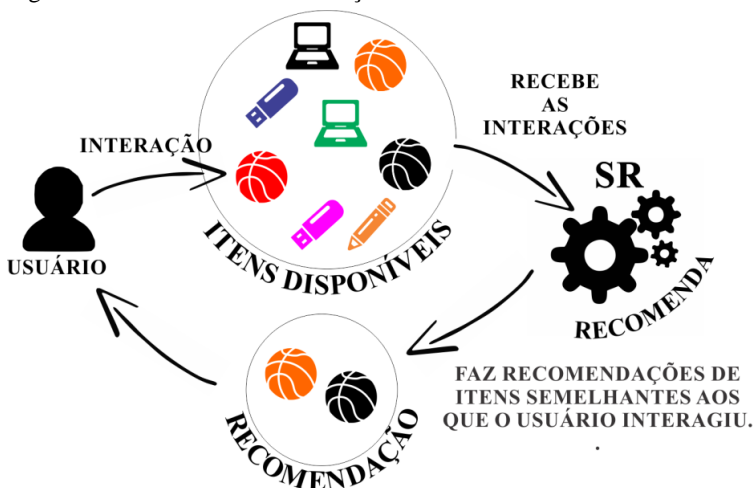
Com intuito de ampliar a visão posta por Burke (2002) e apresentar um cenário mais atual as próximas seções detalham as cinco abordagens citadas pelo autor além das abordagens baseada em comunidade, híbrida e baseada em crítica.

2.2.1 Baseada em Conteúdo

Esta técnica faz uma análise do perfil do usuário e compara com as características dos itens disponíveis. Como exemplo, pode-se citar artigos científicos, em que a partir de um artigo lido é possível sugerir outras leituras comparando as palavras contidas no documento e fazer recomendações semelhantes (CAZELLA; NUNES; REATEGUI, 2010).

O foco desta técnica está na análise de itens a fim de oferecer sugestões que possam interessar a determinado usuário. Funciona de maneira mais confiável quando o item dispõe de uma boa descrição textual. Um exemplo seria a sugestão de filmes, pois em geral possuem sinopse, autor, atores, roteiristas, gênero, ano, entre outras informações relevantes. Assim é possível, por exemplo, selecionar filmes com o mesmo ator, a partir de um interesse que o usuário demonstrou anteriormente (CAZELLA; NUNES; REATEGUI, 2010). A Figura 3 demonstra o processo de recomendação baseada em conteúdo.

Figura 3 - Processo de recomendação baseada em conteúdo.



Fonte: Autor.

Apesar de sua grande utilização esta abordagem possui algumas limitações (ADOMAVICIU; TUZHILIN, 2005), entre elas, a limitação na análise de conteúdo e a superespecialização, detalhadas na seção 2.4.

Além das limitações citadas por Kim et al. (2006) pode-se acrescentar mais duas. A primeira é a dificuldade de fazer recomendações pontuais, pois todas as recomendações devem ser embasadas nas interações e perfil do usuário. A segunda se refere a dificuldade de usuários novos receberem recomendações úteis devido ao fato de seu perfil ser pouco relevante. Porém, a medida que novas interações ocorrem o perfil vai se moldando e tende a se tornar mais robusto melhorando o processo de sugestão.

2.2.2 Filtragem Colaborativa

Esta atualmente é a abordagem mais utilizada, e em conjunto com a baseada em conteúdo são as abordagens mais referenciadas em pesquisas na área. A Filtragem Colaborativa se diferencia da filtragem baseada em conteúdo exatamente por não exigir a compreensão ou reconhecimento do conteúdo dos itens (CAZELLA; NUNES; REATEGUI, 2010). A capacidade de realizar recomendação sem a necessidade de conhecer o conteúdo dos itens pode ser considerada a maior vantagem desta abordagem, permitindo assim, ser aplicada em itens complexos como músicas e filmes (BURKE, 2002).

Esta abordagem tem como premissa que usuários com gostos comuns no passado tendem a repetir determinados padrões e realizar escolhas semelhantes no futuro (HDIOUD; FRIKH; OUHBI, 2012). Adomavicius e Tuzhilin (2005) afirmam que usuários semelhantes provavelmente irão possuir preferência pelos mesmos itens. Diante deste pressuposto é possível afirmar que o perfil do usuário é fundamental para o processo de recomendação. Através de sua análise é possível obter as interações dos usuários com os itens. Segundo Burke (2002), um perfil utilizado na filtragem colaborativa consiste em um vetor de itens e as interações que o usuário realizou, sendo que este vetor tende a crescer a medida que o usuário vai interagindo com o sistema.

Apesar do perfil do usuário sempre ser um vetor de itens e interações, os valores destas interações podem ser distintos, por exemplo, um valor indicando a relevância de determinada avaliação realizada pelo usuário para determinado item.

Esta técnica realiza sugestões com base em usuários com padrões semelhantes em relação a seus gostos, utilizando cálculos como

distância euclidiana, correlação de Pearson, entre outros. Através da aplicação do cálculo escolhido obtém-se a escala de similaridade (SEGARAN, 2008).

A Tabela 1 exemplifica as interações entre usuários e itens para serem utilizadas pela abordagem de filtragem colaborativa. Por exemplo, para recomendar um produto ao Usuário₆ o sistema irá procurar outros usuários com hábitos de semelhantes. A semelhança é obtida através da análise das interações que o usuário faz. Neste exemplo, é possível perceber que o Usuário₆ interagiu com o Item₁ e o Item₆. A partir desta informação é possível encontrar usuários com interesses semelhantes, neste caso o Usuário₁ e o Usuário₄.

Tabela 1 - Relação de interações entre usuários e itens.

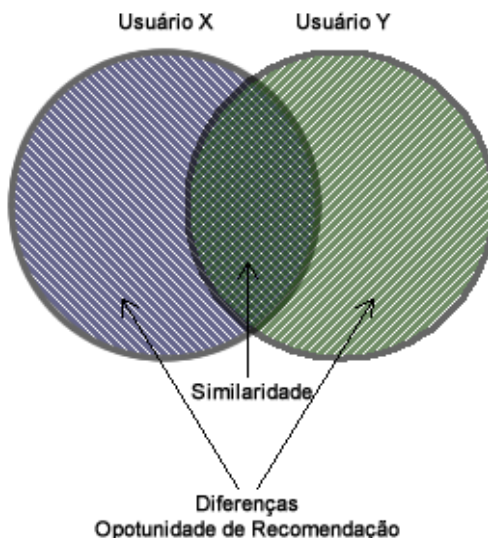
Usuários	Item ₁	Item ₂	Item ₃	Item ₄	Item ₅	Item ₆
Usuário₁	X		X			X
Usuário ₂				X	X	
Usuário ₃	X			X		
Usuário₄		X			X	
Usuário ₅				X	X	
Usuário₆	?	X			?	X

Fonte: Adaptado de Cazella , Nunes e Reategui (2010).

Como pode ser observado o Usuário₁ e o Usuário₄ possuem interesses comuns ao Usuário₆ (Item₂ e Item₆), e possuem interação com itens que o Usuário₆ ainda não interagiu. Sendo assim, a partir desta diferença serão realizadas as recomendações, neste caso o Item₁ e Item₅.

A abordagem analisa dois conjuntos de itens que representam as preferências de usuários distintos e, caso haja uma intersecção, é possível afirmar que os usuários possuem preferências semelhantes permitindo realizar recomendações a partir da diferença obtida entre os dois conjuntos. A Figura 4 ilustra este processo.

Figura 4 - Similaridade, diferenças e oportunidades.



Fonte: Adaptado de Jones (2013).

O sucesso desta abordagem deve-se ao fato de que de modo geral quanto mais similares forem os usuários, mais relevantes serão as recomendações. As principais medidas de similaridade serão apresentadas na seção 2.3.

2.2.3 Demográfica

A abordagem demográfica recomenda itens de usuários demograficamente semelhantes. A análise de usuários pode ser realizada através do gênero, idade, etnia, profissões, nacionalidade, idiomas, entre outros (TINTAREV; MASTHOFF, 2011; PICAULT et al., 2011). Considerar aspectos demográficos representou uma importante evolução para os sistemas de recomendações atuais. Diante desta visão Picault et al. (2011) afirmam que o entendimento destes fatores permite o desenvolvimento de modelos que adequem-se de maneira mais conforme a esta abordagem.

A partir da análise demográfica é possível observar comportamentos comuns, como por exemplo, pode-se afirmar que jovens entre 20 e 25 anos, que residem no sul tem preferência por determinado estilo musical. Essa generalização pode trazer bons

resultados, apesar de produzirem recomendações genéricas que não se aprofundam individualmente no perfil do usuário. O Quadro 2 uma visão geral desta abordagem.

Quadro 2 - Visão geral da abordagem demográfica.

Abordagem	Conteúdo
Descrição	Usuários com atributos demográficos semelhantes são combinados, e a partir disso recomenda itens que são preferidos pelos usuários semelhantes (com base em dados do usuário em vez de classificações).
Vantagens	<ul style="list-style-type: none"> - Sem problema de <i>cold-start</i> (este problema é detalhado na seção 2.4). - Independente de domínio. - Descobertas interessantes ao acaso.
Desvantagens	<ul style="list-style-type: none"> - Dificuldade na obtenção de informação, uma vez que depende da vontade do usuário e em caso de omissão ou erro o resultado será influenciado. - Insuficiência de informação. - Recomenda apenas itens de gostos populares.

Fonte: Adaptado de Drachsler, Hummel e Koper (2007).

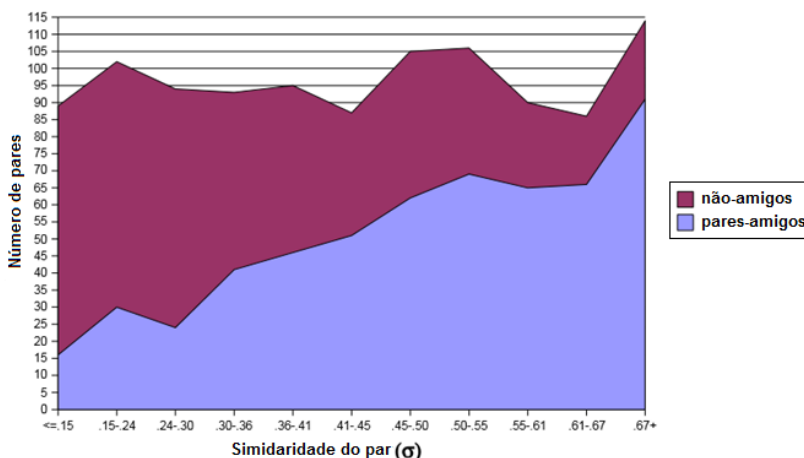
2.2.4 Baseada em Comunidade

Comumente utilizado em redes sociais esta abordagem faz sugestões a partir de preferências de usuários amigos, levando em consideração que o usuário considera uma recomendação de um amigo mais relevante do que a de um anônimo. Este tipo de recomendação é recente e foco de diversos estudos (RICCI; ROKACH; SHAPIRA, 2011).

Apesar de ser semelhante a filtragem colaborativa esta abordagem limita-se a encontrar usuário semelhantes apenas na comunidade que o usuário pertence, diferentemente da filtragem colaborativa que aplica as medidas de similaridade a todos os usuários do sistema.

O foco desta abordagem em um primeiro momento pode ser interpretado como uma limitação na busca por usuários semelhantes. Contudo, Groh e Ehmig (2007) afirmam que usuários mais semelhantes tendem a pertencer a mesma comunidade. O Gráfico 1 demonstra esta afirmação expondo que quanto maior o grau de similaridade, maior a chance do par de usuários pertencerem ao mesmo grupo.

Gráfico 1 - Comparação entre pares de amigos e não amigos.



Fonte: Adaptado de Groh e Ehmig (2007).

A análise de comunidades vem crescendo com as redes sociais e busca relações entre pessoas ou grupos, tais como: equipes, organizações, amigos, familiares, profissões, entre outros (GROH; EHMIG, 2007).

2.2.5 Híbrida

Sistemas de recomendação híbridos combinam duas ou mais técnicas visando obter uma melhor precisão. Esta combinação tenta mitigar os pontos fracos de uma técnica fazendo uma complementação com outras (BURKE, 2002). Normalmente, uma abordagem mais tradicional como a filtragem colaborativa é combinada com outra (BURKE, 2002), sendo atualmente muito comum a utilização da filtragem colaborativa conjuntamente com a baseada em conteúdo (MIDDLETON; ROURE; SHADBOLT, 2009).

Burke (2002) apresenta as principais estratégias utilizadas em sistemas de recomendação híbridos:

- **Ponderadas:** são utilizadas diversas abordagens para se obter os itens que deverão ser recomendados. Nesse processo, cada abordagem pode ter um peso distinto, implicando em um maior ou menor impacto na decisão final;

- **Alternada:** o sistema utiliza a abordagem que mais se adequa a determinada situação. Assim, é possível para o sistema contornar os problemas clássicos com a alternância das abordagens;
- **Mista:** recomendações de diferentes tipos são aplicadas retornando um conjunto de recomendações oriundas de abordagens distintas;
- **Combinação de características:** reúne dados oriundos de diversas bases que são processados por apenas uma abordagem de recomendação;
- **Cascata:** composta por fases, esta estratégia aplica várias abordagens a fim de aprimorar os resultados obtidos.

As pontuações (ou votos) de diversas técnicas de recomendação são combinadas em conjunto para produzir uma única recomendação.

2.2.6 Baseada em Conhecimento

A abordagem de recomendação baseada em conhecimento é uma técnica recente. Estes sistemas utilizam Ontologias e fazem sugestões a partir do conhecimento de determinado domínio (CHEN et al., 2012). A utilização de ontologias conjuntamente com outras abordagens como filtragem colaborativa e baseada em conteúdo tem se apresentado como promissora (MIDDLETON; ROURE; SHADBOLT, 2009).

Segundo Burke (2002), a capacidade de realizar inferências para gerar recomendações pode ser feita por qualquer outra técnica. A capacidade de conhecer um determinado item e as necessidades de um usuário e assim, raciocinar e encontrar uma sugestão adequada é o que difere esta abordagem das demais. A capacidade de conhecer itens e usuário é possível graças à utilização de ontologias que provêem uma estrutura semântica. O Quadro 3 apresenta uma visão geral dos sistemas baseados em conhecimento. No contexto deste trabalho vale ressaltar a necessidade da engenharia do conhecimento para a construção destes sistemas.

Quadro 3 - Visão geral da abordagem baseada em conhecimento.

Abordagem	Conteúdo
Descrição	Sugere itens com base em inferências sobre as necessidades e preferências do usuário.
Vantagens	<ul style="list-style-type: none"> - Pode mapear as necessidades do usuário para os produtos; - Não necessita de <i>ramp-up</i>; - Sensíveis às mudanças de preferências; - Permite também avaliar itens utilizando características externas a ele, tais como: confiabilidade de vendedores e disponibilidade de produtos.
Desvantagens	<ul style="list-style-type: none"> - Engenharia do Conhecimento necessária, ou seja, uma dependência do especialista de domínio; - Recomendações estáticas (não há aprendizado).

Fonte: Adaptado de Burke (2002).

2.2.7 Baseada em Utilidade

Esta técnica utiliza preferências explicitadas pelo usuário para encontrar a melhor opção. Na aplicação desta técnica é fundamental que o usuário conheça algumas características do item que procura, diferentemente de outras técnicas apresentadas (HUANG, 2011).

Diante do Quadro 4 é possível perceber algumas semelhanças entre a abordagem baseada em utilidade e a baseada em conhecimento. Pode-se considerar que a abordagem em questão é uma derivação da baseada em conhecimento, tanto que Felfernig et al. (2011) afirmam que muitas vezes este tipo de abordagem é categorizada como baseada em conhecimento.

Quadro 4 - Visão geral da abordagem baseada em utilidade.

Abordagem	Conteúdo
Descrição	Sugere itens com base em inferências sobre as necessidades e preferências do usuário.
Vantagens	<ul style="list-style-type: none"> - Não necessita de <i>ramp-up</i>; - Sensíveis às mudanças de preferências; - Permite também avaliar itens utilizando características externas a ele, tais como: confiabilidade de vendedores e disponibilidade de produtos.
Desvantagens	<ul style="list-style-type: none"> - O usuário deve informar a função de utilidade; - Recomendações estáticas (não há aprendizado).

Fonte: Adaptado de Burke (2002).

2.2.8 Baseada em Crítica

As recomendações são feitas com base nos *feedbacks* do usuário. Primeiramente, o usuário realiza busca por um item, o sistema retorna os resultados para que o usuário faça os *feedbacks* criticando as características dos itens. A partir da análise das críticas o sistema realiza novas sugestões. Este ciclo se repete até que o usuário encontre um item que atenda suas expectativas (CHEN; PU, 2012).

Na última década diversas propostas e evoluções foram elaboradas nas recomendações baseadas em crítica. O empenho em aprimorar esta abordagem justifica-se pelo seu potencial em fazer recomendações com qualidade.

2.2.9 Outras Abordagens

Além das abordagens de recomendações citadas anteriormente há diversos estudos que aplicam técnicas de Inteligência Artificial (IA) para auxiliar no processo de recomendação. A utilização de IA no processo de recomendação não se configura como uma abordagem diferente, mas sim como técnicas que podem ser utilizadas para gerar recomendações nas abordagens já existentes, tais como, baseada em conteúdo, crítica, utilidade, conhecimento, comunidade, demográfica, filtragem colaborativa, etc.

Dentre as técnicas de IA utilizadas no processo de recomendação o raciocínio baseados em casos e as redes neurais ganham destaque, e são utilizadas em diversos estudos recentes como os de Gatzoura e Sánchez-marrè (2015), Tintarev e Masthoff (2011), Amatriain et al. (2011).

Em geral as técnicas de IA vêm sendo empregadas com intuito de amenizar alguns problemas dos Sistemas de Recomendação (seção 2.4) ou melhorar sua precisão. Um exemplo desta possibilidade é proposta por Gatzoura e Sánchez-marrè (2015), onde afirmam que o raciocínio baseado em casos pode gerar recomendações de itens com base em um conjunto de regras extraídas de um corpus de dados. Gatzoura e Sánchez-marrè (2015) utilizam a técnica de IA com intuito de amenizar o problema de novo item, novo usuário e superespecialização, problemas comuns dos Sistemas de Recomendação.

Além do raciocínio baseado em casos Amatriain et al. (2011) afirmam que também é possível utilizar as técnicas de árvores de

decisão, regras de classificação, redes bayesianas, redes neurais. Esta gama de possibilidades pode ser utilizada conjuntamente contribuindo para a construção de sistemas mais robustos e que mitiguem as deficiências das abordagens de recomendação existentes.

2.3 MEDIDAS DE SIMILARIDADE

Na literatura existem muitas equações capazes de aferir a similaridade entre itens ou usuários. A escolha da equação deve ser realizada através da análise das características do problema e do contexto em que está inserido, sendo que não existe uma técnica em particular que obtenha resultados mais precisos em todos os domínios (HUANG, 2008).

Calcular a similaridade é fundamental durante o processo de recomendação, pois através dos resultados é possível definir se os usuários ou itens são semelhantes ou não (VALOIS; OLIVEIRA, 2011; AL-SHAMRI, 2014). Diante da importância associada às métricas de similaridade existem diversos estudos que avaliam as mesmas. No domínio dos Sistemas de Recomendação as métricas mais utilizadas atualmente são: Distância Euclidiana (AMATRIAIN ET AL., 2011; SGERARAM, 2008), Correlação de Pearson (AL-SHAMRI, 2014; AMATRIAIN et al., 2011; SEGARAM, 2008; TORRES, 2004), Cosseno (AL-SHAMRI, 2014; AMATRIAIN 2011; TORRES 2004) e Coeficiente de Jaccard (AL-SHAMR, 2014; AMATRIAIN ET AL., 2011).

Nas próximas seções serão apresentadas as quatro medidas citadas, com o desenvolvimento de suas respectivas equações. Os vetores utilizados para o desenvolvimento das equações podem ser observados na Tabela 2, que contém dados hipotéticos sobre cinco usuários e três itens, com as respectivas avaliações.

Tabela 2 - Usuários, itens e suas avaliações.

Usuário	Item 1	Item 2	Item 3
José	5	1	4
Mário	4	3	2
Ana	5	2	3
João	2	3	1
Mara	1	4	5

Fonte: Autor.

2.3.1 Distância Euclidiana

A distância euclidiana pode ser considerada uma forma simples e comum de calcular a similaridade entre vetores (AMATRIAIN et al., 2011; SEGARAN 2008). Através da aplicação deste cálculo é possível obter a distância geométrica entre dois conjuntos em um plano multidimensional. Como resultado obtém-se um número real e não negativo, sendo que, quanto maior o resultado maior a distância.

$$d = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

Desenvolvendo a fórmula a partir dos vetores Mara = [1, 4, 5] e João = [2, 3, 1] tem-se:

$$\sqrt{(1 - 2)^2 + (4 - 3)^2 + (5 - 1)^2} = 4.24$$

Ao aplicar a fórmula da distância euclidiana constrói-se uma matriz NxN que apresenta distância entre todos os usuários apresentados a partir dos vetores contidos na Tabela 3.

Tabela 3 - Similaridades dos usuários obtidas através da distância euclidiana.

	José	Mário	Ana	João	Mara
José	0	3	2.82	4.69	5.09
Mário	3	0	1.73	2.23	4.35
Ana	2.82	1.73	0	3.74	4.89
João	4.69	2.23	3.74	0	4.24
Mara	5.09	4.35	4.89	4.24	0

Fonte: Autor.

Neste contexto os valores obtidos demonstram a distância entre dois usuários, ou seja, o quão diferentes são os usuários. Porém segundo Segaram (2008) é possível normalizar os valores utilizando a seguinte equação:

$$s = \frac{1}{d + 1}$$

Após a normalização tem-se uma matriz com valores entre [0, 1], sendo que quanto mais próximo a 1 maior a semelhança entre duas pessoas. Ao final deste processo é produzida uma matriz que demonstra a similaridade entre usuários (Tabela 4).

Tabela 4 - Similaridades normalizadas dos usuários obtidas através da distância euclidiana.

	José	Mário	Ana	João	Mara
José	1,00	0,25	0,26	0,17	0,16
Mário	0,25	1,00	0,36	0,31	0,18
Ana	0,26	0,36	1,00	0,21	0,17
João	0,17	0,31	0,21	1,00	0,19
Mara	0,16	0,18	0,17	0,19	1,00

Fonte: Autor.

2.3.2 Correlação de Pearson

A correlação de Person é a métrica mais utilizada para medir a similaridade em Sistemas de Recomendação (AMATRIAIN et al., 2011). Neste contexto, a correlação de Pearson mede a força do relacionamento entre dois perfis de usuários. O resultado da equação varia entre [-1;1], em que 1 representa usuários idênticos e -1 usuários totalmente opostos. A correlação de Pearson é calculada de acordo com a seguinte equação:

$$w_{a,u} = \frac{\sum_{i=1}^m (r_{a,i} - \bar{r}_a) * (r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i=1}^m (r_{a,i} - \bar{r}_a)^2} \sqrt{\sum_{i=1}^m (r_{u,i} - \bar{r}_u)^2}}$$

Na equação apresentada $w_{a,u}$ representa a correlação entre o usuário a e determinado usuário u . A avaliação que usuário a promoveu

ao o item i é representada como $r_{a,i}$, e \bar{r}_a é a média de todas as avaliações do usuário a . A equação abaixo exemplifica a correlação de Pearson aplicada em dois vetores: José = [5, 1, 4] e Ana = [5, 2, 3].

$$\begin{aligned}
 w_{José,Ana} &= \frac{\sum_{i=1}^m (r_{José,i} - \bar{r}_{José}) * (r_{Ana,i} - \bar{r}_{Ana})}{\sqrt{\sum_{i=1}^m (r_{José,i} - \bar{r}_{José})^2} \sqrt{\sum_{i=1}^m (r_{Ana,i} - \bar{r}_{Ana})^2}} \\
 &= \frac{(5-3.3)(5-3.3) + (1-3.3)(2-3.3) + (4-3.3)(3-3.3)}{\sqrt{(5-3.3)^2 + (1-3.3)^2 + (4-3.3)^2} * \sqrt{(5-3.3)^2 + (2-3.3)^2 + (3-3.3)^2}} = \\
 &= \frac{2.89 + 2.99 + (-0.21)}{\sqrt{8.67} * \sqrt{4.67}} = 0.89
 \end{aligned}$$

Após a aplicação da equação é possível obter a força de correlação; o exemplo desenvolvido demonstra a correlação entre os usuários José e Ana. Após calcular a correlação de todos os usuários utilizando os valores apresentados na Tabela 2 tem-se a matriz de correlação que pode ser visualizada na Tabela 5.

Tabela 5 - Similaridades dos usuários obtidas através da correlação de Pearson.

	José	Mário	Ana	João	Mara
José	1	0,24	0,89	-0,72	-0,5
Mário	0,24	1	0,65	0,5	-0,96
Ana	0,89	0,65	1	-0,32	-0,83
João	-0,72	0,5	-0,32	1	-0,24
Mara	-0,5	-0,96	-0,83	-0,24	1

Fonte: Autor.

Segundo Segaram (2008) a correlação de Pearson é uma equação de similaridade mais sofisticada e exata se comparada a distância Euclidiana. Outro ponto fundamental é que esta métrica oferece uma correção para a inflação de graduação, ou seja, mesmo que o usuário a possua a tendência a atribuir notas mais altas que o usuário b , se eles possuírem preferências semelhantes a força de correlação pode ser alta. A fim de exemplificar este fenômeno podemos apresentar dois vetores: $a = [5, 4, 2, 4]$ e $b = [4, 3, 2, 3]$. Pode-se perceber que o usuário a costuma atribuir notas mais altas, já o usuário b é mais crítico. Mesmo

com esta diferença suas preferências são similares e neste caso o resultado da correlação seria 0.98, sendo considerada elevada.

2.3.4 Cosseno

Esta métrica pode ser utilizada para medir o cosseno do ângulo entre vetores n -dimensionais (TORRES, 2004). Neste contexto n representa o número de itens nos vetores, e cada vetor representa um perfil de usuário com suas respectivas interações. A fórmula do cosseno pode ser apresentada da seguinte forma:

$$\cos(\vec{a}, \vec{u}) = \frac{\vec{a} \cdot \vec{u}}{\|\vec{a}\|_2 \cdot \|\vec{u}\|_2}$$

Na equação apresentada \vec{a} e \vec{u} representam, respectivamente, os vetores dos usuários a e u . O símbolo $\|\vec{a}\|$ indica a norma de \vec{a} que deve ser calculada conforme o exemplo a seguir, considerando os vetores dos usuários João = [2, 3, 1] e Mara = [1, 4, 5].

$$\begin{aligned} \cos_{\text{João}, \text{Mara}} &= \frac{\sum_{i=1}^m (r_{\text{João}, i} * r_{\text{Mara}, i})}{\sqrt{\sum_{i=1}^m (r_{\text{João}, i})^2} \sqrt{\sum_{i=1}^m (r_{\text{Mara}, i})^2}} \\ &= \frac{(2 * 1) + (3 * 4) + (1 * 5)}{\sqrt{(2^2 + 3^2 + 1^2)} \sqrt{(1^2 + 4^2 + 5^2)}} = 0.78 \end{aligned}$$

A aplicação do cálculo apresentado irá retornar um valor entre [0, 1], em que 0 indica que não há nenhuma similaridade e quanto mais próximo a 1 mais semelhantes são os vetores. A

Tabela 6 demonstra o resultado da aplicação desta equação.

Tabela 6 - Similaridades dos usuários obtidas através do cosseno.

	José	Mário	Ana	João	Mara
José	1.00	0.88	0.97	0.70	0.69
Mário	0.88	1.00	0.96	0.96	0.74
Ana	0.97	0.96	1.00	0.94	0.74
João	0.70	0.96	0.94	1.00	0.78
Mara	0.69	0.74	0.74	0.78	1.00

Fonte: Autor.

2.3.5 Coeficiente de Jaccard

Mede a similaridade entre dois conjuntos. No domínio dos sistemas de recomendação esta métrica é comumente utilizada para calcular a similaridade entre dois itens comparando suas propriedades (SEGARAM, 2008). O coeficiente de Jaccard também pode ser utilizado para calcular a similaridade entre dois usuários, porém não considera o valor da avaliação, e sim apenas se o usuário a interagiu com o item i . Este coeficiente é adequado em situações onde o valor das avaliações seja binário, 0 ou 1 (AMATRIAIN ET AL., 2011).

Diante de dois vetores $a = [\text{Item 2}, \text{Item 13}, \text{Item 19}, \text{Item 22}]$ e $b = [\text{Item 1}, \text{Item 2}, \text{Item 10}, \text{Item 13}, \text{Item 22}, \text{Item 27}]$ que representam usuários e suas interações é possível medir sua similaridade a partir da seguinte equação:

$$J = \frac{N_c}{N_a + N_b - N_c}$$

onde c refere-se à intersecção entre os vetores a e b . N_a representa o número de itens avaliados por a , N_b ao número de itens avaliados por b e N_c ao número de itens de c (itens avaliados por a e b).

$$J = \frac{3}{4 + 6 - 3} = 0.42$$

2.4 PROBLEMAS DOS SISTEMAS DE RECOMENDAÇÃO

Este tópico apresenta os principais problemas que atingem os Sistemas de Recomendação. Cada problema afeta de forma diferente as abordagens apresentadas, fazendo com que alguns sistemas utilizem abordagens híbridas com o intuito de mitigar estes desafios. Entre os problemas citam-se:

- **Novo Usuário:** um novo usuário com nenhuma ou pouca interação com o sistema dificulta o processo de recomendação, pois não tem seu perfil definido ou é pouco representativo (CAZELLA; NUNES; REATEGUI, 2010);
- **Novo Item:** a inserção de um novo item, sem interações, faz com que ele não seja recomendado em técnicas de filtragem

colaborativa (CAZELLA; NUNES; REATEGUI, 2010). A medida que o item vai sofrendo interações ele tende a ser recomendado mais frequentemente. Problemas de novo usuário ou novo item muitas vezes são denominados de *ramp-up* ou *cold start*;

- **Esparsidade:** em situações onde há uma grande quantidade de itens e/ou poucos usuários, torna-se difícil encontrar similaridades (VALOIS; OLIVEIRA; 2011). Torres (2004) exemplifica afirmando que em uma empresa como a Amazon® que possui mais de 3 milhões de itens, seria necessário que um usuário consumisse 3.000 itens para abranger apenas 0,1% das opções disponíveis. Este é um problema muito comum e ocorre frequentemente nos Sistemas de Recomendação (BAMBINI; CREMONESI; TURRIN, 2011), fazendo com que em alguns casos gerem recomendações inadequadas (FERDAOUS; BOUCHRA; BRAHIM, 2012);
- **Serendipity (importantes descobertas efetuadas por acaso):** Um sistema preciso, que recomenda somente *best-sellers*, possui menos valor que um sistema que gera recomendações aparentemente inexplicáveis, mas relevantes para alguns usuários. Da mesma forma, um sistema que recomenda itens raros, mas que não são apreciados pelos usuários, não terá uma boa aceitação (TORRES, 2004);
- **Superespecialização do usuário:** a medida que o usuário vai construindo seu perfil através de interações com itens muito semelhantes o sistema tende a gerar recomendações similares. Um usuário que só seleciona filmes de comédia provavelmente no futuro irá receber apenas recomendações deste gênero (ADOMAVICIUS; TUZHILIN, 2005);
- **Tipo do arquivo:** a estrutura e as propriedades dos itens também podem influenciar na qualidade das recomendações. Itens como sons, filmes e imagens não permitem a análise do conteúdo, apresentando realmente um grande problema principalmente para abordagens baseadas em conteúdo (TORRES, 2003).

2.5 CONSIDERAÇÕES FINAIS

Este capítulo apresentou uma visão geral sobre os conceitos fundamentais referentes aos Sistemas de Recomendação. Em seguida foram abordados os benefícios obtidos através da utilização destes sistemas. Após a conceitualização sobre o tema foram abordados aspectos mais aprofundados como as abordagens mais utilizadas para gerar recomendação, bem como as principais medidas de similaridade que dão suporte a tarefa de recomendação. Por fim o capítulo abordou os problemas mais comuns encontrados na área dos Sistemas de Recomendações.

3 REPRESENTAÇÃO DO CONHECIMENTO

A Representação de Conhecimento é uma área da IA, que busca modelar o conhecimento de modo que possibilite a atuação de agentes sobre uma base de conhecimento, permitindo o raciocínio e a tomada de decisão. Segundo Oliveira e Carvalho (2008) “a representação do conhecimento pode ser definida como um conjunto de convenções sintáticas e semânticas que torna possível descrever coisas”.

Através da Representação do Conhecimento é possível conceber representações de objetos e suas relações diante de um contexto previamente delimitado. No âmbito da Ciência da Computação as representações são utilizadas na implementação de estruturas computáveis (CAMPOS, 2004). Segundo Almeida, Souza e Bax (2003) o processo de representação do conhecimento é de fundamental importância, sendo aplicado na maioria dos campos científicos.

A representação do conhecimento é uma forma de criar um modelo formal do conhecimento de determinado domínio com auxílio de um especialista, de modo que possa ser utilizado por um programa de computador. Existe uma gama de estruturas que viabilizam a formalização e organização de conhecimento, sendo que, atualmente as Ontologias tem ganhado destaque neste meio (CASTILHO; LOPES; TACLA, 2008).

Sendo assim, a representação de conhecimento sob a forma de Ontologias possibilita o entendimento mais amplo de determinado domínio aumentando a capacidade de prover informações e de apoiar a tomada de decisão em determinado contexto (SEWALD JUNIOR, 2013).

3.1 ONTOLOGIAS

O termo Ontologia atualmente é utilizado por diversas comunidades (GUARINO, OBERLE, STAAB, 2009). Apesar de originário da Filosofia o termo Ontologia, ainda na década de 60, começou a ser empregado para referenciar um artefato de *software* ou linguagem formal no contexto da Representação do Conhecimento (OLIVEIRA, 2014).

Segundo Gizzard et al. (2009), quando inserido no contexto computacional o termo tem sido utilizado como um “artefato concreto de engenharia, projetado com um propósito específico [...]”, e também para modelar um domínio de conhecimento específico construído através de uma linguagem formal.

Segundo Freitas (2003), no âmbito da Inteligência Artificial a palavra Ontologia “[...] pode ser interpretada como o conjunto de entidades com suas relações, restrições, axiomas e vocabulário”.

Gruber (1993) apresenta um conceito amplamente aceito na literatura, sendo que o mesmo define Ontologia como “uma especificação explícita de uma conceitualização”. A evolução desse conceito foi proposta por Studer, Benjamins e Fensel (1998) em que eles declaram que “uma ontologia é uma especificação formal e explícita de uma conceitualização compartilhada”. Os requisitos deste conceito podem ser esclarecidos da seguinte maneira (BREITMAN, 2006; FREITAS, 2003):

- **Conceitualização:** representa um modelo abstrato de determinada área de conhecimento, onde os conceitos relevantes são identificados;
- **Explícita:** os elementos e suas restrições estão claramente definidos (conceitos, instâncias, relações, restrições e axiomas);
- **Formal:** estrutura bem definida, tornando possível processamento automático da Ontologia (compreensível para agentes humanos ou computacionais);
- **Compartilhada:** a Ontologia utiliza conhecimento consensual em sua modelagem, ou seja, aceito por um grupo de pessoas.

As Ontologias dispõem de cinco tipos de componentes para formalizar o conhecimento, sendo: classes, relações, funções, axiomas e

instâncias (GRUBER, 1993). Gómez-Pérez (1999) descreve os componentes apresentados por Gruber (1993), sendo eles:

- Classes são usadas em um sentido amplo. Elas podem ser abstratas ou concretas, elementares ou compostas, reais ou fictícias. Comumente, as classes são organizadas em forma de hierarquia, formando assim uma taxonomia. Por exemplo, a classe “Humano” é uma subclasse de “Mamífero”;
- Relações tem a finalidade de representar algum tipo de interação entre duas classes de um domínio. Por exemplo, uma relação “*eh_discente*” pode ocorrer entre as classes “Pessoa” e “Universidade”.
- Função é um caso especial de relacionamento em que um conjunto de elementos tem uma única relação com outro elemento. Um exemplo de função é “*ser_pai*”, onde a classe “Homem” e a classe “Mulher” se relacionam com uma classe “Pessoa”.
- Axiomas são verdades absolutas, imutáveis. Por exemplo, é possível afirmar que todo “Humano” é “Mamífero”;
- Instâncias representam elementos específicos pertencentes a uma classe. A pessoa identificada por “Paulo” é uma instância da classe “Humano”.

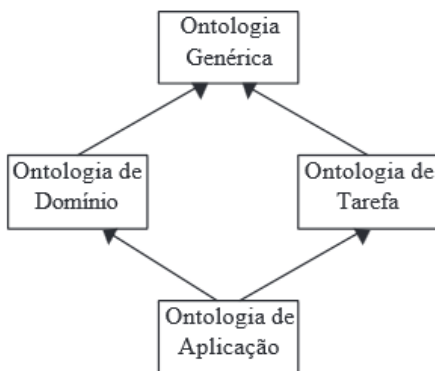
As Ontologias podem ser classificadas em diversos grupos, sendo essa realizada de acordo com sua generalidade. Existem diversas formas de divisão de Ontologias na literatura. Guarino (1998) divide as ontologias em quatro grupos:

- **Ontologias de nível superior:** descrevem conceitos de uso geral, tais como espaço, tempo e eventos. Tal característica possibilita sua utilização para a confecção de novas ontologias independente do domínio (BREITMAN, 2006);
- **Ontologias de domínio:** descrevem domínio específico de uma área genérica, como, por exemplo, uma Ontologia sobre família;
- **Ontologias de tarefas:** descrevem conceitos referentes a uma tarefa ou atividade;
- **Ontologias de aplicação:** descrevem conceitos mais específicos, que geralmente correspondem, de maneira geral, a papéis desempenhados por entidades do domínio durante a

execução de uma tarefa (BREITMAN, CASANOVA, TRUSZKOWSKI, 2007).

A classificação de Guarino (1998) é amplamente aceita e referenciada. A Figura 5 ilustra a percepção do autor quanto ao assunto.

Figura 5 - Tipos de ontologias.



Fonte: Adaptado de Guarino (1998).

Além da classificação proposta por Guarino (1998), Gómez-Pérez, López e Corcho (2002) apresentam uma proposta em que classificam as ontologias através de dois aspectos, expressividade e formalismo, declarando da seguinte maneira:

- **Lightweight:** são ontologias compostas por classes, relações e propriedades, ou seja, não utilizam axiomas e restrições. A modelagem de ontologias consideradas *Lightweight* é adequada quando um nível de expressividade alto não é necessário. Estas características fazem com que o processo de raciocínio realizado sobre este tipo de ontologia seja limitado.
- **Heavyweight:** são ontologias que buscam representar o domínio da forma mais fidedigna possível e permitir uma maior capacidade de raciocínio, fazendo uso constante de axiomas e restrições.

Além das classificações já expostas, Almeida e Bax (2003) apresentam uma coleção de outras classificações, bem como os aspectos que nortearam a classificação e seus autores. Esta sistematização

desenvolvida por Almeida e Bax (2003) pode ser observada no Quadro 5.

Quadro 5 - Proposta de classificação de Ontologias.

Abordagem	Classificação	Descrição
Quanto à função Vanwelkenhuyse e Ikeda (1995)	Ontologias de domínio	Reutilizáveis no domínio; fornecem vocabulário sobre conceitos, seus relacionamentos, sobre atividades e regras que os governam.
	Ontologias de tarefa	Fornecer um vocabulário sistematizado de termos, especificando tarefas que pode ou não estar no mesmo domínio.
	Ontologias gerais	Incluem um vocabulário relacionado a coisas, eventos, tempo, espaço, casualidade, comportamento, funções etc.
Quanto ao grau de formalismo Uschold e Gruninger (1996)	Ontologias altamente informais	Expressa livremente em linguagem natural.
	Ontologias sem-informais	Expressa em linguagem natural de forma restrita e estruturada.
	Ontologias semiformais	Expressa em uma linguagem artificial definida formalmente.
	Ontologia rigorosamente formal	Os termos são definidos com semântica formal, teoremas e provas.
Quanto à aplicação Jasper e Uschold (1999)	Ontologias de autoria neutra	Um aplicativo é escrito em uma única língua e depois convertido para o uso em diversos sistemas, reutilizando-se as informações.
	Ontologias como especificação	Cria-se uma ontologia para um domínio, a qual é usada para a documentação e manutenção no desenvolvimento de softwares.
	Ontologias de acesso comum à informação	Possibilitam a interoperabilidade e compartilhamento de termos quando estes são utilizados em mais de um domínio.
Quanto à estrutura Haav e Lubi	Ontologias de alto nível	Descrevem conceitos gerais relacionados a todos os elementos da ontologia (espaço, tempo, matéria, objeto, eventos, ação, etc.) os quais são independentes do problema ou domínio.
	Ontologias de domínio	Descrevem o vocabulário relacionado a um domínio, como por exemplo,

(2001)		medicina ou automóveis.
	Ontologias de tarefa	Descrevem uma tarefa ou atividade, como, por exemplo, diagnósticos ou compras, mediante a inserção de termos especializados da ontologia.
Quanto ao conteúdo Van-Heijst, Schreiber e Wielinga (2002)	Ontologias terminológicas	Especificam termos que serão usados para representar o conhecimento de um domínio, por exemplo, os léxicos, listas de palavras em uma linguagem que representam um vocabulário junto com algum conhecimento de como cada palavra é usada (HIRST, 2009).
	Ontologias de informação	Especificam a estrutura de registros de banco de dados (por exemplo, os esquemas de banco de dados).
	Ontologias de modelagem do conhecimento	Especificam conceitualizações do conhecimento; possuem uma estrutura interna semanticamente rica e são refinadas para o uso do domínio do conhecimento que descrevem.
	Ontologias de aplicação	Contêm as definições necessárias para modelar o conhecimento em uma aplicação.
	Ontologias de domínio	Expressam conceitualizações que são específicas para um determinado domínio do conhecimento.
	Ontologias genéricas	Similares às ontologias de domínio, mas os conceitos que as definem são considerados genéricos e comuns a vários campos de conhecimento.
	Ontologias de representação	Explicam as conceitualizações que fundamentam os formalismos de representação do conhecimento

Fonte: Almeida e Bax (2003).

Estando ciente das diversas classificações existentes na literatura e visando manter claro em que contexto emprega-se a classificação da Ontologia, esta dissertação baseia-se na classificação proposta por Guarino (1998).

Como mencionado anteriormente a utilização de Ontologias tem como principal objetivo a representação do conhecimento, podendo ser aplicadas em uma fonte de dados previamente existente com o objetivo de melhorar a sua descrição semântica. Segundo Ceci (2010) “Essa aplicação possibilita uma compreensão de um domínio compartilhado entre pessoas e sistemas, adicionando estruturas semânticas a uma fonte de dados e desenvolvendo um intercâmbio de informações”. Além das já citadas às ontologias podem trazer diversos benefícios em uma vasta gama de utilização.

Segundo é possível encontrar na literatura diversas propostas para aplicação em diferentes áreas de conhecimento. Schulz et al., (2009) afirmam que na Biomedicina as ontologias podem atuar, dentre outras áreas na integração, recuperação ou interoperabilidade, pois atualmente os sistemas básicos de terminologia e classificação são inadequados sob estes aspectos.

Outro importante campo que constantemente utiliza as ontologias para aprimorar seus processos é a Gestão do Conhecimento. Nesta área facilitam e suportam atividades de organizações do conhecimento, como: aquisição do conhecimento, além da representação, manutenção e uso do conhecimento.

Atualmente os Sistemas de Recomendação também vem incorporando ontologias, visando adicionar semântica e contornar os problemas clássicos que os atingem (MIDDLETON; ROURE; SHADBOLT, 2009).

A partir deste contexto fica evidente a evolução e a popularização das Ontologias. Além das áreas de conhecimentos já citadas elas atuam frequentemente no comércio eletrônico, o processamento de linguagem natural, a recuperação de informação e aplicações educacionais, entre outros (ALMEIDA; BAX, 2003). A capacidade de representar, reutilizar e interoperar conhecimento, e a possibilidade de realizar inferências sobre este conhecimento são características que promovem às ontologias um papel importante em diversas áreas.

3.2 ONTOLOGIAS DE TAREFA

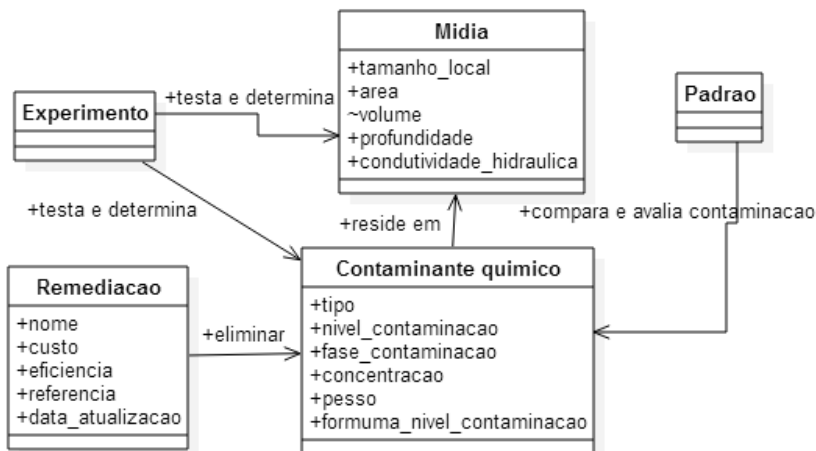
As ontologias de tarefa são voltadas a resolução de problemas específicos de forma genérica, permitindo assim sua reutilização em outros domínios. Estas ontologias fornecem um conjunto de condições, por meio do qual descrevem genericamente como resolver um tipo de problema (MIZOGUCHI; VANWELKENHUYSEN; IKEDA, 1995).

A vantagem da ontologia de tarefa é que ela não só especifica o esqueleto do processo de resolução de problemas, mas também o contexto em que são utilizados os conceitos de domínio (IKEDA; SETA; KAKUSHO, 1998).

Apesar de não ser um conceito novo, essas ontologias são menos maduras que as ontologias de domínio. Trabalhos científicos que abordam este tema são escassos e não descrevem detalhes de como modelar e aplicar este tipo de ontologia. Reforçando estes argumentos, Martins e Falbo (2008) afirmam que o desenvolvimento de ontologias de tarefas, em contraste com o desenvolvimento de ontologias de domínio, não tem maturidade em vários aspectos. Enquanto no desenvolvimento de ontologias de domínio existem várias metodologias, e até certo ponto um acordo sobre o que deve ser capturado em uma ontologia de domínio (conceitos, relações, propriedades e restrições), o mesmo não se aplica ao desenvolvimento de ontologias de tarefa (MARTINS; FALBO, 2008).

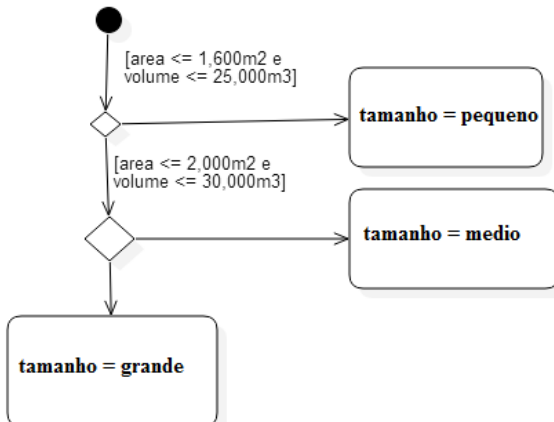
Diante deste contexto trabalhos científicos que abordam ontologias de tarefas utilizam, na grande maioria das vezes, soluções *ad-hoc* e diferentes formas de representação. Wang e Chan (2001) propõem a utilização de UML (*Unified Modeling Language*) para representação de ontologias, utilizando diagramas de classe e de atividade. Neste caso os autores afirmam que a utilização de diagramas de atividade é adequada para representar um processo ou atividade específica (WANG; CHAN, 2001). A Figura 6 e a Figura 7 demonstram a utilização da UML para a representação de uma ontologia de reparação de danos causados por petróleo.

Figura 6 - Fragmento da ontologia de reparação de danos causados por petróleo.



Fonte: Wang e Chan (2001).

Figura 7 - Determinando o tamanho da média (visão da operação).

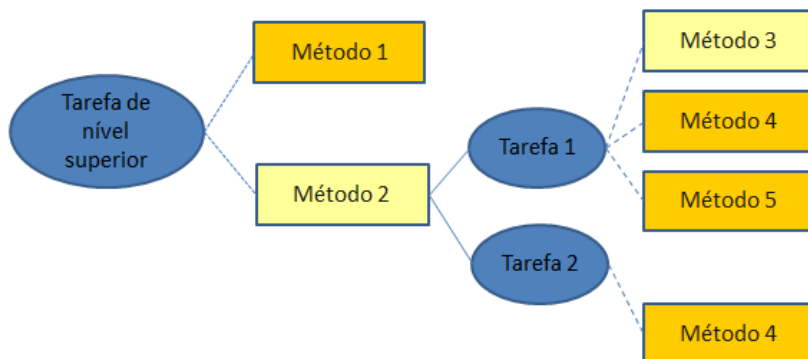


Fonte: Wang e Chan (2001).

Com uma percepção semelhante, Martins (2009) propõe a representação das ontologias através de duas perspectivas: estrutural e comportamental. Para a representação estrutural é proposto a utilização de diagramas de classe (UML), e para a comportamental sugere-se a utilização de uma derivação dos diagramas de caso de uso, atividade e estado, também providos pela UML.

Além da UML como alternativa de representação Buckeridge et al. (2008), propõem uma estrutura que pode ser utilizada para representar tarefas de forma adequada. Na estrutura as tarefas são representadas por elipses, que podem ser compostas por um ou mais métodos, representados por retângulos, tais componentes pode ser analisados na Figura 8.

Figura 8 - Exemplo de árvore de decomposição de tarefa.



Fonte: Adaptado de Buckeridge et al., (2008).

Uma tarefa define o que deve ser realizado. Para chegar ao objetivo da tarefa são executados métodos, estes por sua vez definem como executar determinada tarefa. Há dois tipos de métodos distintos, um método que executa uma tarefa diretamente, chamada de método primitivo, e o chamado de composto que quebra uma tarefa em subtarefas (BUCKERIDGE et al., 2008).

3.2 METODOLOGIAS DE ENGENHARIA DE ONTOLOGIA

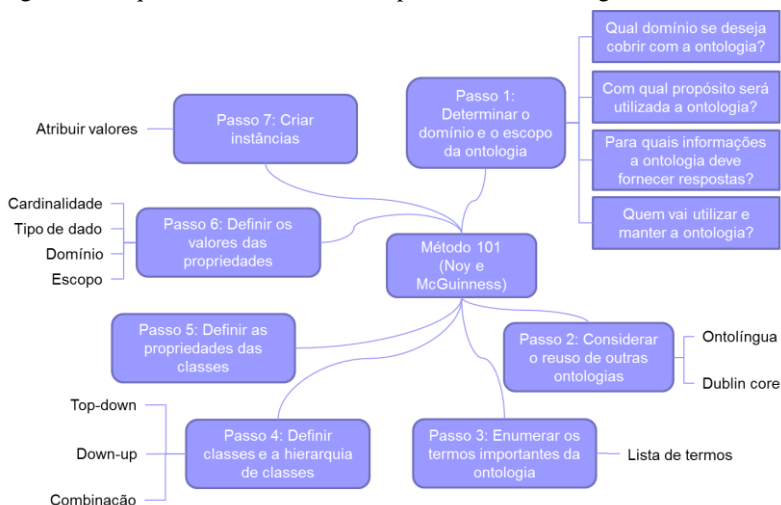
Este capítulo abordará metodologias para o desenvolvimento de ontologias. As metodologias abordadas são: Metodologia 101, Methontology, KACTUS e TOVE, sendo que serão apresentados os principais aspectos bem como as etapas que as compõem.

3.2.1 Metodologia 101

Esta metodologia consiste em um processo iterativo, detalhado por passos que guiam os envolvidos, mesmo que possuam uma grande experiência em Engenharia de Ontologia.

A Metodologia 101 (Ontology Development 101) possui uma estrutura simples e passos diretos e bem detalhados, características que facilitam o processo de modelagem de ontologias e destacam-na como uma metodologia amplamente utilizada. A Figura 9 apresenta o conjunto de passos da Metodologia 101 (NOY; MCGUINNESS, 2001).

Figura 9 - Esquema resumindo com os passos da Metodologia 101.



Fonte: Elaborado a partir do trabalho de Noy e McGuinness (2001).

A *Ontology Development 101* é composta por sete passos iterativos (RAUTENBERG et al., 2008) sendo eles:

- **Determinar o escopo da ontologia:** este passo determina questões fundamentais que limitarão o escopo da ontologia, tais como o domínio, quem irá utiliza-la, as questões que a ontologia deverá responder e que manterá a ontologia seu desenvolvimento.
- **Considerar o reuso:** avalia a possibilidade de utilizar ontologias já existentes, podendo ser utilizadas de forma parcial ou integral. Existem diversas bibliotecas de ontologias reusáveis que pode ser exploradas para este fim.
- **Listar termos importantes:** neste passo os termos usados na ontologia são listados.

- **Definir classes:** definem-se as classes da ontologia e sua hierarquia. É possível seguir a abordagem *top-down*, *bottom-up* ou uma combinação das duas anteriores.
- **Definir restrições das propriedades (Relações):** neste passo são adicionadas as propriedades de modo que a ontologia consiga responder as perguntas de competência.
- **Criar instâncias:** por fim são criadas instâncias das classes.

3.2.2 Metodologia METHONTOLOGY

Desenvolvida pelo Laboratório de Inteligência Artificial da Universidade Politécnica de Madrid, da Espanha, em 1997, a METHONTOLOGY é uma metodologia para construção de novas ontologias iniciadas a partir do zero, ou por utilização da reengenharia sobre ontologias já existentes (FERNADEZ-LOPEZ et al., 1997).

A METHONTOLOGY é uma metodologia cíclica e evolutiva, sendo que a construção de ontologias envolve alguns estágios: especificação, aquisição do conhecimento, conceitualização, formalização, integração, implementação, avaliação, documentação e manutenção (MATTOS; SIMÕES; FARIAS, 2007).

3.2.3 Metodologia KACTUS

O método KACTUS (*modelling Knowledge About Complex Technical systems for multiple Use*) foi concebido através do projeto europeu *Esprit Kactus* que objetivava a construção de ontologias no domínio de redes elétricas. O foco do projeto era investigar a viabilidade da reutilização de conhecimento em sistemas técnicos complexos e o papel da ontologia para apoiá-los (SILVA; SOUZA; BARCELLOS, 2008).

Seguindo os padrões do KACTUS a criação de uma ontologia esta diretamente relacionada ao desenvolvimento de uma aplicação. Assim quando uma nova aplicação é construída a estrutura responsável por representar o conhecimento, ou seja, uma ontologia, também é construída (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ, 2002). Os passos executados na construção de uma ontologia são os seguintes:

- **Especificação da aplicação:** determina um contexto da aplicação e proporciona uma visão dos componentes que aplicação tenta modelar.

- **Projeto Preliminar:** determinam-se conceitos relevantes de alto nível da ontologia através da análise da etapa anterior. Também é viável a análise de ontologias já desenvolvidas, que após um refinamento e adequação são passíveis de utilização.
- **Refinamento e estruturação da ontologia:** com intuito de chegar ao projeto definitivo a ontologia é refinada de modo a obter um artefato mais coerente possível.

3.2.4 Metodologia TOVE

Proposta por Michael Gruninger e Mark Fox em 1995 (GRUNINGER; FOX, 1995) esta metodologia foi denominada TOVE em alusão ao projeto que deu origem a mesma: *Toronto Virtual Enterprise*. Esta metodologia se caracteriza por transformar cenários informais, descritos em linguagem natural em modelos de linguagem formal, tornando-os assim computáveis (FERNÁNDEZ, 1999). As etapas propostas pela metodologia são as seguintes (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ, 2002; SILVA; SOUZA; BARCELLOS, 2008):

- **Captura dos cenários de motivação:** tem como objetivo identificar problemas no cenário atual ou deficiências das ontologias existentes.
- **Especificação de questões de competência informal:** descreve em linguagem natural questões que a ontologia deve contemplar. As questões são obtidas através da etapa anterior.
- **Especificação da terminologia da ontologia em uma linguagem formal:** nesta etapa são definidos termos que devem ser capazes de responder as questões de competência descritas na etapa anterior. Os termos especificados e organizados em uma taxonomia e suas propriedades são bases para especificar a ontologia em uma linguagem formal.
- **Formulação das questões de competência formais:** as questões de competência são transcritas utilizando lógica de primeira ordem. Neste momento as questões já podem ser definidas em uma linguagem formal.
- **Especificações dos axiomas:** a ontologia é codificada e linguagem lógica e formal, fazendo com que os axiomas restrinjam a interpretação dos termos envolvidos nas questões de competência.

- **Verificação da completude da ontologia:** definir as condições sob quais as soluções para as questões de competência são completas.

3.3 CONSIDERAÇÕES FINAIS

Primeiramente apresentou-se uma introdução sobre a Representação de Conhecimento. Em seguida o capítulo apresentou as Ontologias seus principais conceitos, características e grupos em que podem ser categorizadas. Após isso as ontologias de tarefas foram detalhadas e abordaram-se algumas formas de representação desta categoria de ontologia. Por fim algumas metodologias para a construção de ontologias foram apresentadas e detalhadas.

4 MODELO PROPOSTO

Neste capítulo será detalhado o modelo proposto que consiste em uma ontologia genérica capaz de representar o domínio de Sistemas de Recomendação. Além disso, é proposta uma ontologia de tarefa que possui como objetivo prover suporte a realização de recomendações considerando algumas das abordagens discutidas anteriormente.

A partir do escopo delimitado neste trabalho o início da elaboração do modelo ocorreu pela escolha da metodologia. Diante da importância da modelagem na representação do conhecimento a Engenharia de Ontologia oferece uma série de metodologias para este fim. No presente trabalho a metodologia selecionada foi a *Ontology Development 101*, proposta por Noy e McGuinness (2001). Esta metodologia foi utilizada pela sua simplicidade e pelo detalhamento das etapas, características que se adequam melhor a construção de uma ontologia de pequeno porte, em que não há uma equipe envolvida no desenvolvimento.

O processo de construção e a estrutura da ontologia de domínio serão detalhados nas sessões 4.1 até a 4.7. A sessão 4.8 aborda a ontologia de tarefa, detalhando sua estrutura e suas características.

4.1 DETERMINANDO O ESCOPO DA ONTOLOGIA

Neste passo os princípios para a construção da ontologia são definidos. Para determinar o escopo a etapa é guiada por uma série de perguntas que auxiliam a delinear determinado modelo de interesse (NOY; MCGUINNESS, 2001).

a) Qual é o domínio que a ontologia abrangerá?

O domínio de conhecimento que a ontologia abrangerá será o dos Sistemas de Recomendação.

b) Para que a ontologia será utilizada?

O modelo possui como objetivo a realização de recomendações. Além de tornar viável a tarefa de recomendação, o modelo proposto pode ser reutilizado, permitindo assim, agregar outras abordagens de recomendação, como a baseada em conteúdo, demográfica, entre outras. Outro aspecto relevante é a geração de conhecimento que a Ontologia irá possibilitar quando for aplicada, dando subsídios para a tomada de decisão através da análise do comportamento dos usuários e suas interações.

c) Que tipos de perguntas a ontologia deveria ser capaz de responder?

A ontologia proposta poderá responder o seguinte conjunto de perguntas, também entendidas como questões de competência:

- Qual o grupo (por exemplo, por profissão, por interesses, entre outros) que determinado usuário pertence?
- Quais as categorias preferidas pelo usuário?
- Quais itens sofreram alguma interação por determinado usuário?
- Quais são os itens mais bem avaliados?
- Quais são os itens mais bem avaliados em determinada categoria?
- Quais são os itens que sofreram mais interações?
- Quais são os itens que sofreram mais interações em determinada categoria?
- Quais são os perfis de usuários mais semelhantes a determinado usuário considerando seu perfil?

A lista apresentada expõe as principais questões que a ontologia deverá responder, porém não contempla todo o conjunto de possíveis perguntas estando de acordo com a visão de Noy e McGuinness (2001),

em que afirmam que a lista não precisa ser exaustiva e sim contemplar as principais questões que o modelo irá abranger.

d) Quem irá usar e manter a ontologia?

A ontologia poderá ser aplicada em diversos sistemas que procuram facilitar o processo de escolha por parte dos usuários. Como o tipo de conteúdo recomendado é genérico, a ontologia pode sugerir livros, filmes, artigos, pessoas, viagens, etc. Quanto à manutenção da ontologia, essa deve ser realizada pelos administradores do sistema que a utilizará.

4.2 CONSIDERAR O REUSO

Diante da importância do reuso no contexto de representação do conhecimento e principalmente na elaboração de Ontologias realizou-se uma pesquisa por modelos já existentes que poderiam ser utilizados integral ou parcialmente. A pesquisa abrangeu os principais repositórios de Ontologia, bem como, foram realizadas pesquisas nos buscadores tradicionais.

A pesquisa objetivando identificar ontologias em cinco repositórios específicos utilizando os termos “*recommender*” e “*recommender system*”. O Quadro 6 demonstra os repositórios que a pesquisa abordou.

Quadro 6 - Repositórios de ontologias utilizados na pesquisa.

Nome	URL
BioPortal	http://bioportal.bioontology.org/
prefix.cc	http://prefix.cc/
TONES Project	http://rpc295.cs.man.ac.uk:8080/repository/
Repositório Protégé	http://protegewiki.stanford.edu/wiki/Protege_Ontology_Library#OWL_ontologies
VIVO Ontology Repository	http://swl.slis.indiana.edu/repository/

Fonte: Autor.

Considerando estes repositórios não foram encontrados modelos passíveis de utilização perante o contexto apresentado.

4.3 LISTAR TERMOS IMPORTANTES

Este passo consiste em listar os termos relevantes de maneira simples. A construção desta lista é o ponto de partida, sendo assim, não é necessária uma atenção maior quanto a questões de termos com mesmo conceito, relações entre termos ou propriedades que tais termos podem possuir (NOY; MCGUINNESS, 2001).

Os termos relevantes apresentados no Quadro 7 foram extraídos do Capítulo 2 desta dissertação, que aborda os Sistemas de Recomendação.

Quadro 7 - Lista inicial de termos relevantes.

Avaliação	<i>Like</i>	<i>Rating</i>
Categoria	Nota	Recomendação
Comentário	Perfil do Usuário	Serviço
Compra	Pessoa	Similaridade
Grupo	Predição	Sugestão
Item	Produto	Usuário

Fonte: Autor.

4.4 DEFINIR CLASSES

A partir da análise da lista construída na etapa anterior é possível definir as classes que irão compor a ontologia. Os termos do Quadro 7 identificados como classes são detalhados no Quadro 8.

Quadro 8 - Lista de classes obtidas.

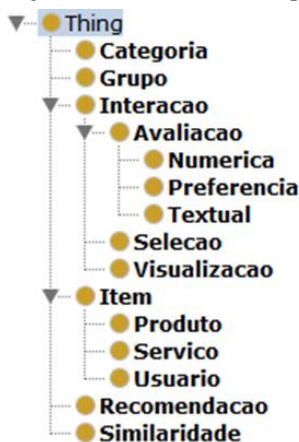
Classe	Descrição
Categoria	Representa a categoria de um item que pode ser. Ex: filmes, séries, documentários.
Grupo	Representa o grupo que um usuário pertence. Os grupos podem variar de acordo com o contexto em que o sistema está inserido, por exemplo, os usuários podem ser classificados através da profissão, interesses, comunidades, entre outros.
Interacao	Classe que representa as interações de um usuário com os itens.
Avaliacao	Subclasse de interação. Representa uma interação que visa avaliar de forma crítica um item.
Preferencia	Adequação do termo <i>Like</i> declarado no Quadro 7. Subclasse de Avaliação. Representa uma avaliação binária, ou seja, se um usuário gostou ou não gostou do item.
Numerica	Adequação do termo <i>Rating</i> . Subclasse de Avaliação. A classe Numerica representa as avaliações através de uma escala predefinida.
Textual	Subclasse de Avaliação. Representa avaliações descritivas dos itens, sendo realizada através de comentários.
Selecao	Subclasse de Interação. Representa uma escolha efetiva de um item. Ex: uma compra, solicitação de amizade, <i>download</i> de um artigo.
Visualização	Subclasse de interação. Representa uma visualização simples de um item.
Item	Representa os produtos, serviços e pessoas passíveis de interações e que serão futuramente recomendados.
Produto	Subclasse de Item. Classe dos bens tangíveis.
Servico	Subclasse de Item. Classe dos bens intangíveis.
Usuario	Subclasse de Item. Classe dos usuários do sistema.
Recomendacao	Classe que define os itens que serão sugeridos a determinado usuário.
Similaridade	Classe que representa o grau de similaridade entre itens.

Fonte: Autor.

Outro aspecto importante é a estruturação hierárquica das classes através da definição de subclasses. Uma subclasse é uma especialização de uma superclasse, por exemplo, a superclasse “ser vivo” tem uma subclasse chamada “humano”. A partir da estrutura deste exemplo é possível afirmar que todo “humano” é um “ser vivo”.

As classes identificadas acima são organizadas de forma hierárquica com classes e subclasses. A Figura 10 apresenta a estrutura obtida após a estruturação das classes, e a inserção no editor de Ontologias Protégé®.

Figura 10 - Estrutura hierárquica das classes do modelo.



Fonte: Autor.

4.5 DEFINIR PROPRIEDADES

Propriedades são relações binárias que podem ser aplicadas entre dois indivíduos ou entre um indivíduo e um valor de dado. Segundo McGuinness, Smith e Welt (2004) propriedades podem ser enquadradas em duas categorias.

- **Propriedades de objeto** (*object properties*): relação entre dois indivíduos. Ex: João -> temSimilaridade-> Paulo.
- **Propriedades de dados** (*datatype properties*): relação entre indivíduos e valores de dados. Ex: João -> temDataNascimento -> “10/12/1987”.

Além do tipo, uma propriedade possui duas características importantes que auxiliam na sua correta representação, *domain* (domínio

ou pertinência) e *range* (alcance, escopo ou classe que preenche ou conecta a propriedade). Por exemplo, uma propriedade “ehDiscente” possuiria como *domain* a classe **Pessoa** e como *range* a classe **Universidade**. Isso implica que a propriedade pertence à classe **Pessoa** e que somente instâncias da classe **Universidade** podem ser atribuídas para ela.

Ontologias em OWL (linguagem utilizada para representar computacionalmente) têm valores de dados definidos pela W3C. Dentre uma lista extensa de valores o Quadro 9 demonstra alguns tipos, que são utilizados no modelo.

Quadro 9 - Alguns tipos primitivos de dados.

Tipo	Descrição
String	É uma cadeia de caracteres com espaço finito.
Boolean	Estado lógico que suporta apenas dois valores [<i>true</i> , <i>false</i>].
DateTime	Representa uma data, sendo composta pelo ano, mês, hora, minuto, segundos, e fuso horário.
Int	Assume um valor de um número inteiro entre -2147483648 e 2147483647.
Double	Representa números fracionados de dupla precisão (64 bits).

Fonte: Adaptado de W3C (2009).

Os tipos de dados demonstrados no Quadro 9 são utilizados para a construção do Quadro 10 que apresenta as propriedades de dado do modelo, onde a coluna “Classe” representa o domínio, a coluna “Propriedade” representa a propriedade de dado e a coluna “Tipo” define o dado que será atribuído. Neste caso, a coluna “Classe” representa o *domain* e a coluna “Tipo” representa o *range*.

Quadro 10 - Propriedade de dado do modelo.

Classe	Propriedade	Tipo
Categoria	descricao_grupo	string
	ID_categoria	int
Grupo	descricao_grupo	string
	ID_grupo	int
Interação	data_interacao	dateTime
	ID_interacao	int
Numerica	valor_avaliacao_numerica	int
Preferencia	valor_avaliacao_preferencia	int
Textual	conteudo_avaliacao	string
	valor_avaliacao_textual	int
Selecao	valor_selecao	int
Visualizacao	valor_visualizacao	dateTime
Item	conteudo_item	string
	ID_item	int
	selecionavel	boolean
	tipo_conteudo	string
Usuario	cep	int
	data_nascimento	dateTime
	meida_avaliacoes	double
	numero_avaliacoes	int
	Nome	string
	Sexo	string
Recomendacao	ID_recomendacao	int
	valor_predicao	double
Similaridade	ID_similaridade	int
	valor_similaridade	double

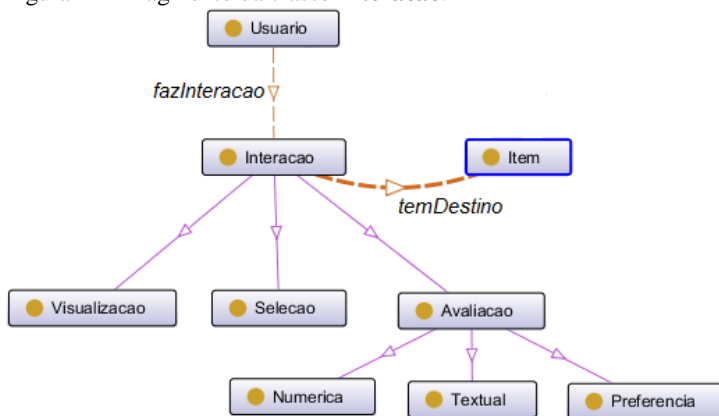
Fonte: Autor.

A seguir serão abordadas as propriedades de objeto, que conecta dois indivíduos. Para melhor entendimento e a visualização das propriedades apresentadas o modelo será dividido em fragmentos.

4.5.1 Fragmento da classe **Interacao**

A Figura 11 demonstra a hierarquia da classe **Interacao** e suas relações com as classes **Usuario** e **Item**.

Figura 11 - Fragmento da classe **Interacao**.



Fonte: Autor.

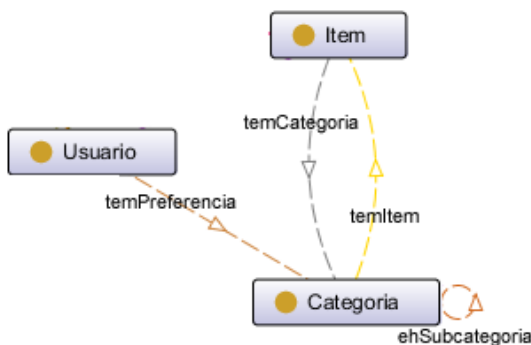
As interações são realizadas por usuários, e tem como alvo os itens (produto, pessoa, ou serviço). A propriedade **fazInteracao** tem como *domain* um usuário e como *range* uma interação, determinando assim qual usuário realizou uma interação. Após registrar quem fez determinada interação é necessário definir para qual item foi destinada a interação, para isso criou-se a propriedade **temDestino** em que as classes **Interacao** e **Item** representam *domain* e *range*, respectivamente.

O fragmento apresentado, responsável pela representação das interações é um ponto fundamental para um sistema de recomendação. A partir da análise das interações é possível determinar o comportamento e os *feedbacks* que os usuários realizam. O conhecimento obtido através desta análise que representa o comportamento de determinado usuário é fundamental para auxiliar no processo de tomada de decisão.

4.5.2 Fragmento da classe **Categoria**

O fragmento apresentado na Figura 12 é responsável por representar as categorias dos itens e as categorias prediletas dos usuários.

Figura 12 - Fragmento da classe **Categoria**.



Fonte: Autor.

A representação permite a criação de uma lista de categorias e através da propriedade **ehSubcategoria** é possível a inclusão de subcategorias, característica que pode ser comumente percebida nos sistemas de recomendações atuais.

Além da hierarquia de categorias é possível definir as categorias preferidas dos usuários através da propriedade **temPreferencia** que conecta usuários e categorias. Existe ainda as propriedades **temCategoria** e **temItem**, que respectivamente associam categorias a um item e determinam quais itens pertencem a determinada categoria.

De maneira geral, a classe **Categoria** permite a organização os itens, além de definir as preferências do usuário. Tais características são utilizadas frequentemente nos sistemas de recomendação.

4.4.4 Fragmento da classe **Grupo**

Este fragmento representa os grupos de usuários (Figura 13). Os usuários podem ser agrupados por profissão, interesses, comunidades, entre outros.

Figura 13 - Fragmento da classe **Grupo**.



Fonte: Autor.

Como pode ser visto na Figura 13 um grupo contém usuários, sendo que, a propriedade de objeto **temUsuario** determina o relacionamento entre os usuários de um grupo. Esta propriedade tem como *domain* a classe **Grupo** e *range* a classe **Usuario**. Também é possível determinar quais grupos um usuário pertence através da propriedade inversa **fazParte**.

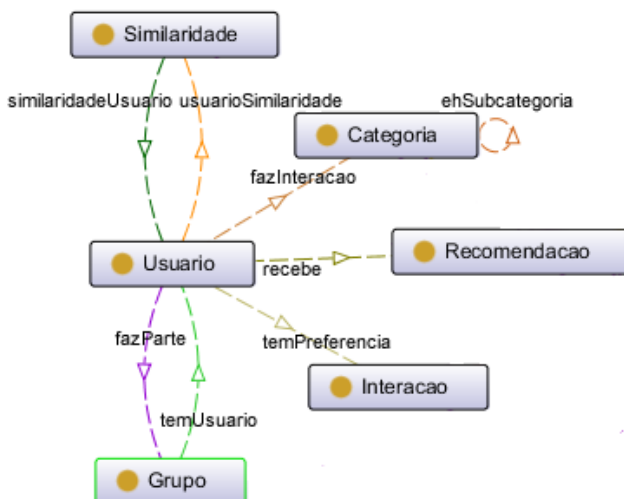
O fragmento dos grupos é simples, porém através de sua análise o perfil dos usuários é possível visualizar comportamentos de grupos distintos. Esta característica torna o conhecimento sobre o comportamento dos usuários muito mais rico e aprofundado.

4.4.5 Fragmento da classe **Usuario**

Os usuários são fundamentais para um sistema de recomendação. A ontologia proposta apresenta uma estrutura para os usuários bem sucinta, porém considerando a premissa de reuso seria possível uma extensão dessa estrutura com intuito de contemplar todos os requisitos existentes no domínio de aplicação.

A classe **Usuario** apresentada na Figura 14 tem interação direta com as classes **Grupo**, **Interacao**, **Categoria**, **Similaridade** e **Recomendacao**. Dentre as classes citadas às duas primeiras já foram abordadas e a classe **Recomendacao** será descrita na próxima seção. Sendo assim, este fragmento contempla a relação entre dois usuários quaisquer indicando o relacionamento com uma instância de similaridade que irá conectá-los.

Figura 14 - Fragmento da classe **Usuario**.



Fonte: Autor.

Analisando a Figura 14 é possível observar as propriedades de objeto entre as classes **Usuario** e **Similaridade**. A propriedade **usuarioSimilaridade** possui como *domain* a classe **Usuario** e a classe **Similaridade** como *range*. Além disso, existe uma propriedade inversa chamada **similaridadeUsuario**. A estrutura proposta é capaz de definir o quão similar dois usuários são conectando determinada instância de usuário a uma instância de similaridade e, esta por sua vez, será conectada a uma instância de usuário.

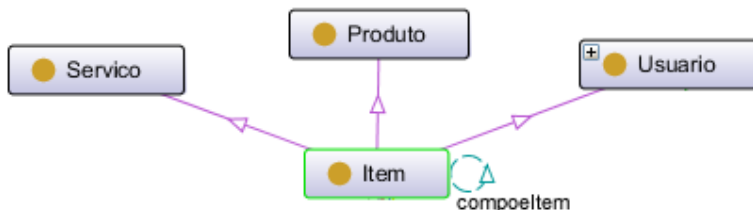
Além da similaridade entre usuários, o perfil do usuário obtido através de suas características é de fundamental importância. No modelo o perfil do usuário se caracteriza por suas interações, preferências e dados demográficos. Cada usuário possui um perfil que tende a se alterar a medida que o usuário interage com o sistema.

Em sumo o fragmento apresentado é capaz de representar o perfil dos usuários através de suas interações, grupos, categorias e similaridade com outros usuários. Através da análise deste perfil é possível gerar recomendações embasando-se em outros usuários com perfis semelhantes.

4.4.5 Fragmento da classe **Item**

Uma recomendação tem como objetivo sugerir um ou mais itens para um usuário. Com a proposta de uma ontologia genérica, ou seja, capaz de recomendar qualquer tipo de item construiu-se a estrutura apresentada na Figura 15.

Figura 15 - Fragmento da classe **Item**.



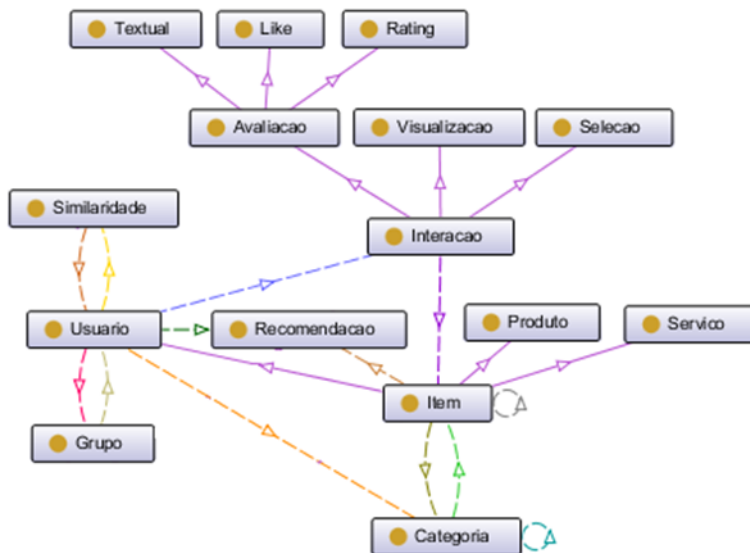
Fonte: Autor.

Primeiramente, após uma análise da literatura foram definidas as subclasses de itens. Deste modo, um item pode ser entendido como um serviço, produto ou usuário. Além das subclasses a propriedade de objeto *compoeItem* é apresentada no fragmento, possuindo como *domain* e *range* a classe **Item**. Esta propriedade possibilita a composição de itens de maneira genérica, bem como, a recomendação de itens com as mais diversas características.

4.4.1 Ontologia de domínio

A partir da aplicação da metodologia apresentada, chegou-se ao seguinte modelo que apresenta a interconexão entre as classes com o objetivo de representar genericamente uma ontologia para o domínio de Sistemas de Recomendação (Figura 16).

Figura 16 - Estrutura da ontologia proposta.



Fonte: Autor.

4.6 DEFINIR RESTRIÇÕES DAS PROPRIEDADES

Para melhor apresentar as restrições do modelo este tópico foi dividido em duas partes, a primeira aborda as propriedades de dado e a segunda as propriedades de objeto.

4.6.1 Propriedades de Dados

Esta etapa detalha os valores e as relações das propriedades apresentadas na etapa anterior, indicando o tipo de valores que irão suportar, o tipo de classe que a compõe e a cardinalidade permitida entre as classes e os valores.

Os tipos de dados definem o tipo de valor de uma propriedade, no caso de uma propriedade simples. O Quadro 9 apresenta os tipos de dados utilizados no modelo e faz uma descrição dos mesmos. Além dos tipos apresentados, o W3C (2009) apresenta a lista completa e suas respectivas descrições.

A escolha do tipo de dados determina o escopo que será empregado à propriedade. Por exemplo, um nome de usuário, do tipo *String*, é capaz de representar uma cadeia de caracteres, tais como:

Marcos, João, Pedro, etc. Propriedades do tipo *dateTime* tem capacidade de representar um momento específico, por exemplo, uma data de nascimento ou o momento em que o usuário fez uma interação. O tipo *boolean* assume apenas dois valores, falso (*false*) ou verdadeiro (*true*) e é utilizado para definir se um item é selecionável ou não. Para representar o valor obtido através das equações de similaridade o tipo *double* é utilizado, pois tem a capacidade de representar números fracionados, ex: 1.7838. Já o tipo *Int* é utilizado em situações onde os números sempre serão inteiros, como: número de avaliações de um determinado usuário, o valor de avaliação (*rating*) que varia de 0 até 5.

Após definir os tipos de dados utilizados, o foco passa a ser a cardinalidade, que neste caso define quantos valores são permitidas para determinada propriedade. A cardinalidade de uma propriedade deve assumir um valor mínimo, um valor exato, ou um valor máximo. O Quadro 11 demonstra os valores de cardinalidade que uma relação pode possuir.

Quadro 11 - Exemplos de cardinalidade mínima, exata e máxima.

Valor	Exemplo
Valor mínimo (Min.)	Um usuário deve ter no mínimo 0 cep. O valor Min. 0 deixa a propriedade como opcional.
Valor exato (Exat.)	Um usuário tem exatamente 1 nome.
Valor Máximo (Max.)	Uma categoria pode ter no máximo uma descrição.

Fonte: Autor.

Definidos os valores possíveis para a cardinalidade aplicou-se o conceito nas propriedades do modelo. O resultado pode ser observado no Quadro 12.

Quadro 12 - Cardinalidade das propriedades do modelo.

Classe	Propriedade	Tipo	Cardinalidade
Categoria	descricao_categoria	<i>String</i>	Exat. 1
Usuario	cep	<i>String</i>	Min. 0 Max. 1
	num_avaliacoes	<i>Int</i>	Exat. 1
	sexo	<i>String</i>	Min. 0 Max. 1
	nome	<i>String</i>	Exat. 1
	data_nascimento	<i>DateTime</i>	Min. 0 Max. 1
Grupo	descricao_grupo	<i>String</i>	Exat. 1
Interacao	data_interacao	<i>DateTime</i>	Exat. 1
Preferencia	valor_avaliacao_preferencia	<i>Int</i>	Exat. 1
Textual	valor_avaliacao_textual	<i>String</i>	Exat. 1
Numerica	valor_avaliacao_numerica	<i>Int</i>	Exat. 1
Selecao	valor_selecao	<i>DateTime</i>	Exat. 1
Visualizacao	valor_visualizacao	<i>DateTime</i>	Exat. 1
Item	conteudo_item	<i>String</i>	Exat. 1
	selecionavel	<i>Boolean</i>	Exat. 1
	tipo_conteudo	<i>String</i>	Exat. 1
Similaridade	valor_similaridade	<i>Double</i>	Exat. 1

Fonte: Autor.

Diante da apresentação das cardinalidade fica evidente que a grande maioria possui o valor Exat. 1. Há também propriedades em que a cardinalidade varia de Min. 0 e Max. 1, fazendo com que o valor torne-se opcional e limite-se no máximo a 1.

4.6.2 Propriedades de Objeto

As propriedades de objeto conectam um indivíduo a outro. Os valores de cardinalidade utilizados são os mesmos já detalhados anteriormente, porém este tipo de propriedade não conecta indivíduos a tipos de dados primitivos.

As propriedades de objeto do modelo, detalhadas a seguir, serão divididas considerando-se o *domain* das classes seguindo a divisão

adotada na Seção 4.5. Sendo assim o Quadro 13 apresenta as restrições de propriedade da classe **Usuario**.

Quadro 13 - Restrições de propriedades da classe **Usuario**.

Propriedade	Cardinalidade	Range
<i>usuarioInteracao</i>	Min. 0	Interacao
<i>fazParte</i>	Min. 0	Grupo
<i>recebeRecomendacao</i>	Min. 0	Recomendacao
<i>temPreferencia</i>	Min. 0	Categoria

Fonte: Autor.

A classe **Usuario** possui cinco propriedades de objeto, sendo que todas possuem a cardinalidade mínima zero. A propriedade *usuarioInteracao* permite conectar diversas interações a um usuário. Esta característica permite que o usuário realize diversas interações ao decorrer do tempo. Há também a possibilidade de determinar as categorias preferidas de um usuário através da propriedade *temPreferencia*, a relação do usuário com n categorias. A propriedade *recebeRecomendacao*, possibilita um usuário receber diversas recomendações ao longo do tempo. Por fim, a propriedade *fazParte*, define que um usuário participe de vários grupos, que podem ser divididos por profissão, interesse, entre outros.

A classe **Item** possui seis propriedades, que são detalhadas na Quadro 14. Assim como a classe **Usuario** as propriedades da classe **Item** possuem cardinalidade mínima zero.

Quadro 14 - Restrições de propriedades da classe **Item**.

Propriedade	Cardinalidade	Range
<i>compoeItem</i>	Min. 0	Item
<i>ehComposto</i>	Min. 0	Item
<i>itemInteracao</i>	Min. 0	Interacao
<i>itemSimilaridade</i>	Min. 0	Similaridade
<i>temCategoria</i>	Min. 0	Categoria

Fonte: Autor.

O modelo proposto utiliza a ideia de que um item é composto por diversos outros itens. A propriedade *compoeItem* determina que um item compoem outros itens; Seguindo a mesma linha, a propriedade *ehComposto* conecta um item a todos que o compõe. A relação entre os

itens e interações é definido com a propriedade *itemInteracao*, que possui cardinalidade mínima zero, possibilitando que um item se conecte a várias interações. A relação entre um item e uma similaridade é representada através da propriedade *itemSimilaridade*, fazendo com que um item possua diversas similaridades. A conexão entre item e suas possíveis categorias é representado pela propriedade *temCategoria*, com cardinalidade mínima zero, permitindo assim que um item não possua nenhuma, uma, ou diversas categorias.

A classe **Categoria** contém apenas duas propriedades de objeto. A propriedade *ehSubCategoria* com cardinalidade extada 1, permite com que uma categoria seja subcategoria de apenas uma categoria. Por outro lado, a segunda propriedade *temSubCategoria*, com cardinalidade mínima zero permite que uma categoria possua diversas subcategorias, como demonstrado no Quadro 15.

Quadro 15 - Restrições de propriedades da classe **Categoria**.

Propriedade	Cardinalidade	Range
<i>ehSubCategoria</i>	Exat. 1	Categoria
<i>temSubCategoria</i>	Min. 0	Categoria

Fonte: Autor.

A classe **Interacao** possui apenas uma propriedade de objeto, sendo esta denominada *interacaoItem* de acordo com a Quadro 16. A propriedade faz com que uma interação se conecte com apenas um item. Sendo assim, é possível que um usuário realize inúmeras interações, porém uma interação sempre estará relacionada a um item específico.

Quadro 16 - Restrição de propriedade da classe **Interacao**.

Propriedade	Cardinalidade	Range
<i>interacaoItem</i>	Exat. 1	Item

Fonte: Autor.

Por fim, tem-se a propriedade da classe **Recomendacao**, chamada de *recomendacaoItem*, que é detalhada no Quadro 17. Esta propriedade permite com que uma recomendação contenha exatamente 1 item. Porém, a propriedade *recebeRecomendacao*, abordada anteriormente permite que um usuário recebe *n* recomendações.

Quadro 17 - Restrição de propriedade da classe **Recomendacao**.

Propriedade	Cardinalidade	Range
<i>recomendacaoItem</i>	Exat. 1	Item

Fonte: Autor.

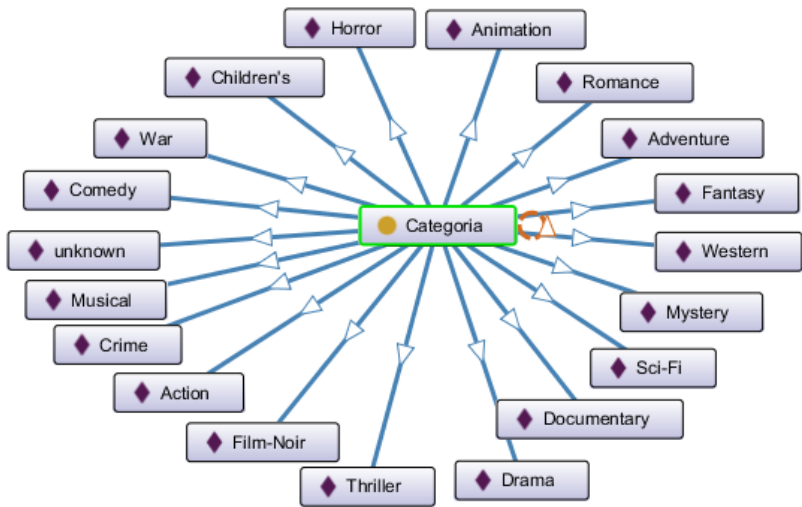
A partir das propriedades de dado e objeto apresentadas é possível obter um maior entendimento da Ontologia. Esta etapa de detalhamento é fundamental no processo de lapidação da Ontologia, uma vez que propriedades bem estruturadas são essenciais para se alcançar um bom nível de formalismo.

4.7 CRIAR INSTÂNCIAS

A última etapa da metodologia 101 é responsável pela criação das instâncias da ontologia. Os indivíduos (instâncias) utilizados no preenchimento da ontologia foram retirados do conjunto de dados apresentado no Capítulo 5.

As primeiras instâncias criadas foram da classe **Categoria**. Apesar do modelo suportar uma estrutura de categorias e subcategorias, foram criadas apenas categorias de mesmo nível. A Figura 17 demonstra 19 categorias que foram instanciadas.

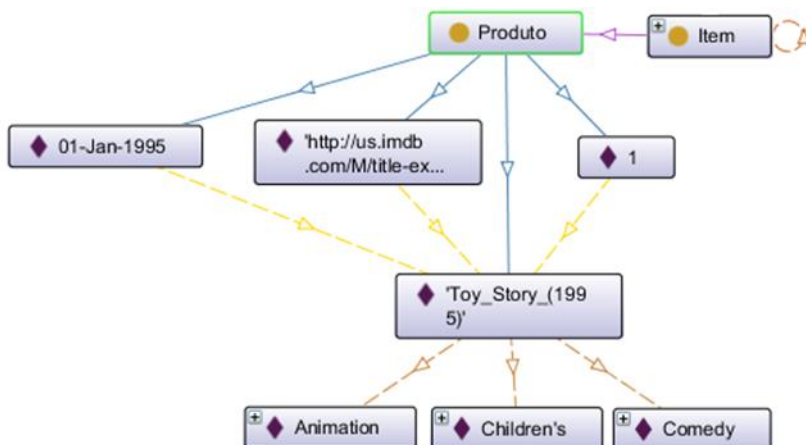
Figura 17 - Instâncias da classe **Categoria**.



Fonte: Autor.

A segunda classe a ser instanciada foi a **Item**. Como há uma grande quantidade de itens a Figura 18 demonstra apenas um exemplo com intuito de facilitar a visualização. Também é possível perceber a relação do item com as categorias abordadas anteriormente. Fica evidente ainda a estrutura de composição de itens, onde o item “Toy Story” é composto por uma data de lançamento, seu endereço no site de crítica IMDB¹, e o identificador do filme.

Figura 18 - Instâncias da classe **Item**.

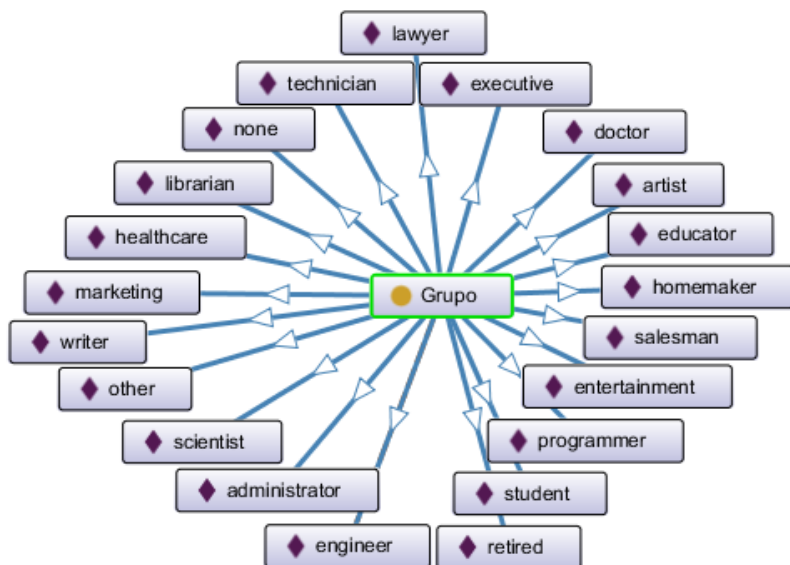


Fonte: Autor.

As instâncias da classe **Grupo** também podem ser vistas na Figura 19. Os 21 grupos instanciados referem-se a profissão dos usuários obtidas a partir do conjunto de dados utilizados.

¹ <http://www.imdb.com/>

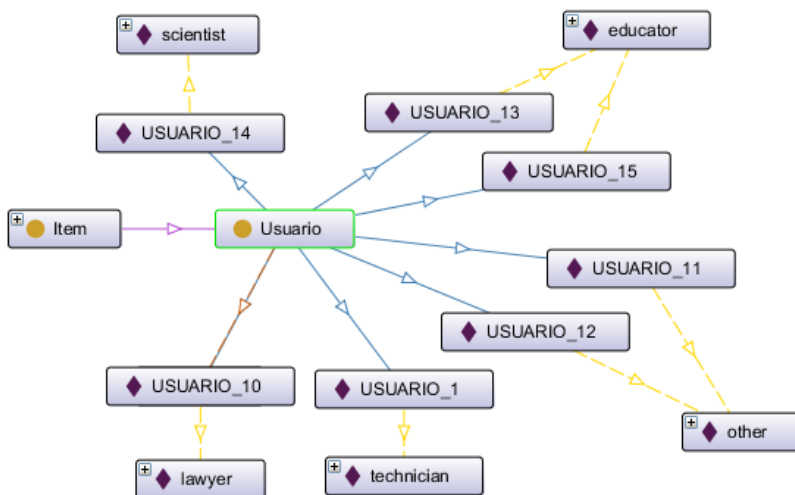
Figura 19 - Instâncias da classe **Grupo**.



Fonte: Autor.

Uma vez que as instâncias da classe **Grupo** tenham sido inseridas, é possível demonstrar o relacionamento entre estas e instâncias da classe **Usuario**. Assim como na demonstração dos itens, a Figura 20 exibe um número limitado de usuários para facilitar a visualização. Também é possível observar o grupo que cada usuário pertence.

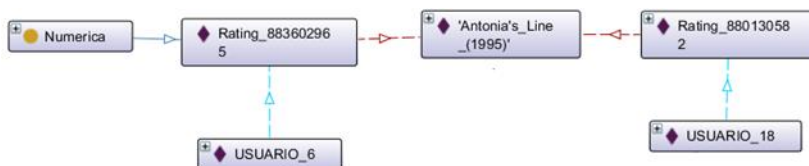
Figura 20 - Instâncias da classe **Usuario**.



Fonte: Autor.

Por fim, foram adicionados os indivíduos da classe **Numerica**, representando as avaliações realizadas previamente pelos usuários, e que estavam no conjunto de dados. A Figura 21 demonstra as avaliações e seu relacionamento com o item a quem se destina e com o usuário que avaliou.

Figura 21 - Instâncias da classe **Numerica**.



Fonte: Autor.

Finalizando este passo conclui-se a aplicação da metodologia. Nesta etapa foram definidas as classes, propriedades e restrições do modelo. Adicionalmente, foram incluídas instâncias para as classes **Categoria**, **Grupo**, **Usuario**, **Item** e **Interacao** (subclasse **Numerica**). Com a inclusão dos indivíduos e da formalização do modelo, torna-se possível aplicar determinada abordagem de recomendação, permitindo

medir a similaridade entre usuários e realizar sugestões, cenário que será detalhado no Capítulo 5.

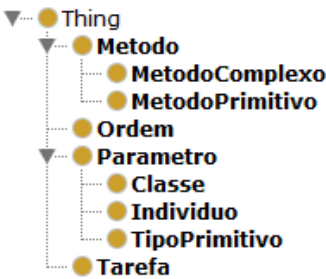
4.8 ONTOLOGIA DE TAREFA

A ontologia de tarefa proposta visa representar o conhecimento sobre as tarefas que devem ser executadas para gerar as recomendações. Esta estrutura foi adicionada, pois a ontologia de domínio de maneira isolada não tem a capacidade de realizar as tarefas necessárias contidas no processo de recomendação.

A falta de metodologias para a construção e o detalhamento deste tipo de ontologia na literatura dificulta seu desenvolvimento. A partir desta realidade a proposta desenvolvida não utilizou nenhuma metodologia, porém foi construída com base na proposta de Buckeridge et al. (2008) apresentada na sessão 3.2.

Após a análise estrutural das tarefas e dos aspectos técnicos envolvidos para sua execução chegou-se a hierarquia de classes ilustrada na Figura 22.

Figura 22 - Estrutura hierárquica das classes da ontologia de tarefa.



Fonte: Autor.

Além da hierarquia apresentada anteriormente o Quadro 18 apresenta as descrições das classes propostas na ontologia.

Quadro 18 - Lista de classes da ontologia de tarefa.

Classe	Descrição
Metodo	Classe que representa a decomposição de tarefas. Podem ser classificados como complexos ou primitivos.
MetodoComplexo	Método que possui decomposição, quebrando uma tarefa em subtarefas.

MetodoPrimitivo	Métodos que não possuem decomposição, sendo constituído por apenas uma etapa.
Ordem	Classe que contem as ordens de execução das tarefas.
Parametro	Representa as possíveis entradas que uma tarefa recebe para ser executada.
Individuo	Parâmetro que recebe uma instância de uma classe.
Classe	Parâmetro que representa uma classe da ontologia
TipoPrimitivo	Parâmetro que recebe um tipo de dado primitivo. Ex: <i>int</i> , <i>string</i> , <i>boolean</i> .
Tarefa	Processo a ser executado através de uma sucessão de passos representados pelos métodos.

Fonte: Autor.

As propriedades de dado das classes que compõem a ontologia de tarefa são descritas no Quadro 19. As propriedades *metodoComplexoNome*, *metodoPrimitivoNome*, *tarefaNome*, representam respectivamente o nome das classes **MetodoComplexo**, **MetodoPrimitivo** e **Tarefa**. A classe **MetodoPrimitivo** possui as propriedades *metodoNomeMetodo* e *metodoNomeClasse* que contêm o nome do método e o nome da classe do algoritmo responsável por executar o método. A propriedade *ordemValor* possui um número inteiro que representa uma ordem para a execução de um método ou tarefa.

Quadro 19 - Lista de propriedades de dado das classes da ontologia de tarefa.

Classe	Propriedade	Tipo
MetodoComplexo	<i>metodoComplexoNome</i>	string
MetodoPrimitivo	<i>metodoPrimitivoNome</i>	string
	<i>metodoNomeClasse</i>	string
	<i>metodoNomeMetodo</i>	string
Ordem	<i>ordemValor</i>	int
Parametro	<i>parametroTemOrdem</i>	int
	<i>parametroTemValor</i>	string
Tarefa	<i>tarefaNome</i>	string

Fonte: Autor.

A partir da descrição das classes da ontologia de tarefa é possível definir a estrutura e o relacionamento entre as mesmas, esta definição é exposta na Figura 23. A classe **Tarefa** se relaciona com a classe **Ordem** através da propriedade *temOrdem* que define a ordem de

execução da mesma. Há também um relacionamento entre **Tarefa** e **Metodo** denominado *temMetodo*, que determina os métodos que compõem determinada tarefa. Por fim, a classe **Tarefa** possui um relacionamento com **Parametro** indicando qual o parâmetro de entrada de determinada tarefa.

4.9 CONSIDERAÇÕES FINAIS

Através do capítulo 4 foram descritas as duas ontologias utilizadas no modelo proposto. Primeiramente definiu-se a ontologia de domínio através da aplicação da metodologia 101 proposta por Noy e McGuinness (2001). No decorrer da execução dos passos da metodologia foram detalhadas características fundamentais como o escopo, perguntas de competência e detalhes estruturais da ontologia.

Por fim, já com a ontologia de domínio especificada iniciou-se a construção da ontologia de tarefa. O desenvolvimento da ontologia de tarefa foi guiado pela ontologia de domínio com intuito de ser utilizada conjuntamente para realizar as recomendações.

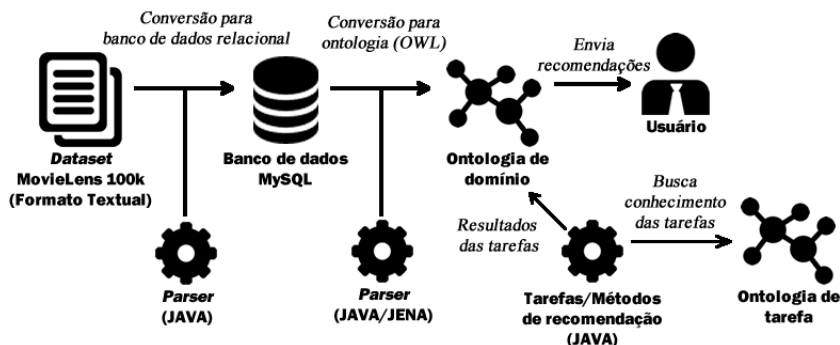
Sendo assim construiu-se uma ontologia capaz de representar o domínio de recomendação (ontologia de domínio) e outra (ontologia de tarefa) com objetivo de gerar recomendações sobre a primeira. A aplicação do modelo e os testes realizados serão detalhados no capítulo 5.

5 APLICAÇÃO DO MODELO E DISCUSSÃO DOS RESULTADOS

Este capítulo descreve a aplicação do modelo proposto e apresenta uma discussão dos resultados obtidos. Entretanto, antes de descrever os resultados serão abordadas algumas questões do desenvolvimento e preparação dos dados, etapas prévias e fundamentais para a aplicação do modelo.

A Figura 25 apresenta uma visão geral do processo realizado desde a coleta e adequação dos dados, até a utilização da ontologia de domínio para prover as respostas às questões de competências, bem como, a utilização da ontologia de tarefa para prover recomendações.

Figura 25 - Descrição detalhada do modelo.



Fonte: Autor.

As seções a seguir detalham do fluxo apresentado na figura. Primeiramente, será detalhado o cenário utilizado para a avaliação do modelo proposto composto por um conjunto de dados de análises de filmes. A partir do conjunto de dados, foram realizadas adequações e o desenvolvimento de um *parser* capaz de transferir os dados em formato original (formato de textual) para um banco de dados estruturado. Isto foi realizado para facilitar a manipulação e entendimento dos dados a serem utilizados no cenário de avaliação.

Um segundo *parser* foi elaborado com o objetivo de transferir os dados estruturados para o modelo semântico (Ontologia de domínio) proposto. Neste passo, as instâncias foram carregadas e as propriedades que as conectam foram definidas.

Em seguida é possível observar um terceiro passo diretamente responsável pela tarefa intensiva em conhecimento de recomendação. A

partir da análise da ontologia de tarefa e da ontologia de domínio as recomendações são elaboradas e persistidas na ontologia de domínio. Para gerar as recomendações utilizou-se a Correlação de Pearson como medida de similaridade, considerando sua precisão e a capacidade de analisar o padrão comportamental do usuário, característica que permite que usuários sejam similares quando possuam comportamentos parecidos e não somente quando compartilham os mesmos valores de avaliação.

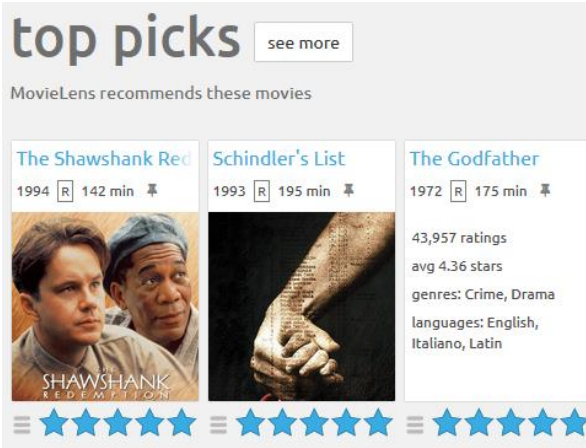
Na etapa final, essas recomendações vão ser apresentadas para o usuário. A apresentação através de um sistema está fora do escopo deste trabalho. Todavia, considerando a ontologia populada com os dados de origem e com as recomendações derivadas a mesma facilmente poderia ser acessada por um sistema visando a disponibilização para um usuário final.

5.1 DESCRIÇÃO DO CENÁRIO DE AVALIAÇÃO

Com o objetivo de avaliar o modelo proposto, composto por uma ontologia de domínio e uma ontologia de tarefa voltadas ao domínio de sistemas de recomendação, utilizou-se um conjunto de dados originado do site *MovieLens*®.

Este *site* é mantido e dirigido pelo *GroupLens Research* da Universidade de Minnesota. O *MovieLens*® usa a abordagem de filtragem colaborativa para fazer recomendações. Além de ser amplamente utilizado com intuito de selecionar filmes, o *MovieLens*® é também foco de estudos na área de Sistemas de Recomendação. A Figura 26 e a Figura 27 obtidas a partir do site do *MovieLens*® demonstram alguns resultados de recomendações efetuadas.

Figura 26 - Site de recomendações de filmes *MovieLens*®.



Fonte: Imagem obtida a partir do site MovieLens®².

Figura 27 - Site de recomendações de filmes MovieLens (Filme *The Godfather*).



Fonte: MovieLens®.

Com o objetivo de fomentar a pesquisa o site MovieLens® através do *GroupLens Research*³ disponibiliza dados reais das interações realizadas por usuários cadastrados em sua base de filmes. Apesar de disponibilizar as bases de dados nenhuma informação que possa

² <https://movielens.org/home>

³ <http://grouplens.org/datasets/movielens/>

explicitamente identificar determinado usuário é provida, ação que visa manter a privacidade.

Atualmente existem três bases disponíveis oriundas do site *MovieLens*®. A Tabela 7 detalha as bases de dados acessíveis.

Tabela 7 - Bases de dados disponibilizadas pelo GroupLens.

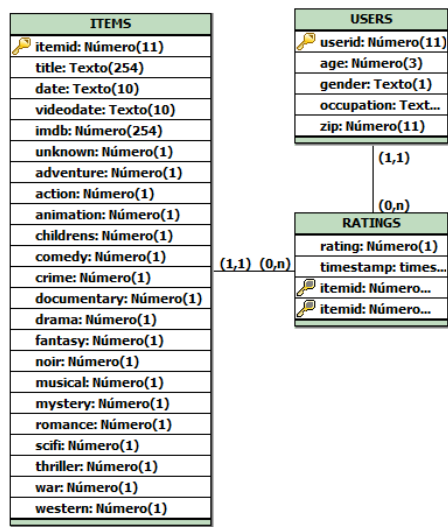
Nome da base	Número de usuários	Número de avaliações	Número de itens
MovieLens 100k	1.000	100.000	1.700
MovieLens 1M	6.000	1.000.000	6.000
MovieLens 10M	72.000	10.000.000	72.000

Fonte: Autor.

A partir das possibilidades apresentadas optou-se pelo menor conjunto de dados, reconhecendo as limitações computacionais em processar estes grandes conjuntos de dados.

As bases de dados disponíveis possuem algumas características distintas no que se refere a quantidade de informação disponível sendo que o modelo da base de dados *MovieLens*® 100k possui uma estrutura que permite a construção de tabelas. Como mencionado anteriormente, essas informações são disponibilizadas originalmente na forma textual. Para tal, foi desenvolvido um *parser* permitindo assim a criação e população do modelo apresentado na Figura 28.

Figura 28 - Estrutura da base MovieLens 100k.



Fonte: Autor.

A tabela *ITEMS* possui 23 campos e representa a informações sobre os 1700 itens. A descrição dos campos é a seguinte:

- **itemid**: identificador, composto por um número sequencial e inteiro;
- **title**: título do item;
- **date**: data de lançamento;
- **videodate**: campo não utilizado;
- **imdb**: link para o site da Amazon® onde constam informações e críticas sobre determinado filme.

Além destes campos, a tabela possui mais 18 campos que correspondem às categorias do filme, sendo elas: *unknown*, *action*, *adventure*, *animation*, *children's*, *comedy*, *crime*, *documentary*, *drama*, *fantasy*, *film-noir*, *horror*, *musical*, *mystery*, *romance*, *sci-fi*, *thriller*, *war*, *western*. A Tabela 8 demonstra a estrutura de um filme cadastrado.

Tabela 8 - Estrutura de um registro (filme) contido na tabela *ITEMS*.

Campo	Valor
-------	-------

<i>itemid</i>	1
<i>title</i>	Toy Story (1995)
<i>date</i>	01-Jan-1995
<i>videodate</i>	null
<i>imdb</i>	http://us.imdb.com/M/title-exact?Toy%20Story%20(1995)
<i>unknown</i>	0
<i>action</i>	0
<i>adventure</i>	0
<i>animation</i>	1
<i>children's</i>	1
<i>comed</i>	1
<i>crime</i>	0
<i>documentary</i>	0
<i>drama</i>	0
<i>fantasy</i>	0
<i>film-noir</i>	0
<i>horror</i>	0
<i>musical</i>	0
<i>mystery</i>	0
<i>romance</i>	0
<i>sci-fi</i>	0
<i>thriller</i>	0
<i>war</i>	0
<i>western</i>	0

Fonte: Autor.

Após a descrição da tabela *ITEMS* pode-se analisar a estrutura da tabela *USERS* (Tabela 9), que contém a identificação e as informações demográficas dos usuários. Esta tabela é composta por cinco campos, sendo eles:

- **userid:** identificador do usuário, composto por um número sequencial e inteiro;
- **age:** número inteiro que representa a idade do usuário;
- **gender:** gênero, M para masculino (*male*) e F para feminino (*female*);
- **occupation:** representa a profissão do usuário, tendo como opções: *administrator, artist, doctor, educator, engineer, entertainment, executive, healthcare, homemaker, lawyer, librarian, marketing, none, other, programmer, retired, salesman, scientist, student, technician, writer*;

- **zip:** identificador postal.

Tabela 9 - Estrutura de um registro (usuário) contido na tabela USERS.

Campo	Valor
<i>userid</i>	7
<i>age</i>	57
<i>gender</i>	M
<i>occupation</i>	administrator
<i>zip</i>	91344

Fonte: Autor.

A última tabela que compõe o modelo de dados utilizado é denominada *RATINGS* (Tabela 10) e possui as informações sobre a avaliação dos filmes. Esta tabela é composta por quatro campos:

- **rating:** valor da avaliação realizada pelo usuário. O valor varia entre 0 e 5;
- **timestamp:** registro do momento em que a avaliação foi realizada;
- **itemid:** identificador do item avaliado;
- **userid:** identificador do usuário que avaliou o item.

Tabela 10 - Estrutura de um registro (avaliação) contido na tabela RATINGS.

Campo	Valor
<i>Rating</i>	5
<i>Timestamp</i>	877717833
<i>Itemid</i>	1
<i>Userid</i>	57

Fonte: Autor.

Apesar de a estrutura ser simples esta representa de maneira satisfatória um conjunto de itens, usuários e suas interações. A partir destes dados serão realizados todos os testes e avaliações desenvolvidas neste capítulo.

5.2 POPULANDO A ONTOLOGIA

Primeiramente, antes de adicionar as instâncias foram necessárias algumas adequações para que o conjunto de dados se tornasse compatível com o modelo proposto. Tais adequações foram necessárias para tornar o tipo dos atributos oriundos do conjunto de dados compatíveis com a ontologia. As adequações realizadas podem ser observadas no Quadro 20 e Quadro 21.

Quadro 20 - Adequação do conjunto de dados para o modelo proposto (data de nascimento).

Informação	Descrição
Como era <i>(dataset):</i>	O conjunto de dados não informa a data de nascimento, apenas a idade, composta por um número inteiro. Ex: 29.
Como deveria ser:	No modelo a data de nascimento assume o formato de <i>dateTime</i> , que tem a seguinte estrutura “2002-10-10+05:00” (ano-mês-dia-fuso horário).
Como ficou:	Através da idade informada no conjunto de dados, é possível estimar (aproximado) o ano de nascimento. O dia e mês sempre irão assumir a data de 1 de janeiro. Ex: “1986-01-01+5:00” para um usuário com 29 anos.

Fonte: Autor.

Quadro 21 - Adequação do conjunto de dados para o modelo proposto (Categoria preferida).

Informação	Descrição
Como era (dataset):	No conjunto de dados não há informação quanto às categorias preferidas do usuário.
Como deveria ser:	Um usuário pode definir as categorias preferidas.
Como ficou:	Após uma análise do perfil dos usuários definiu-se que a categoria preferida é aquela que o usuário mais avaliou. Em caso de empate, ambas são aceitas.

Fonte: Autor.

Além das mudanças apresentadas foi necessário fazer um recorte do conjunto de dados. A redução da base foi necessária por dois motivos. O primeiro refere-se a instabilidade e as inconsistências apresentadas pelo editor de ontologias utilizado no desenvolvimento do modelo, o Protégé®. O segundo é a incapacidade de carregar todo o conjunto de dados para a ontologia, devido a limitações de recursos como a Memória RAM. A Tabela 11 demonstra a situação do conjunto de dados antes e depois do recorte realizado na base.

Tabela 11 - Comparação entre os conjuntos de dados, completo e reduzido.

Conjunto de Dados	Número de usuários	Número de avaliações	Número de itens
Original (MovieLens 100k)	1.000	100.000	1.700
<i>Dataset</i> utilizado (Reduzido)	29	1.712	1.664

Fonte: Autor.

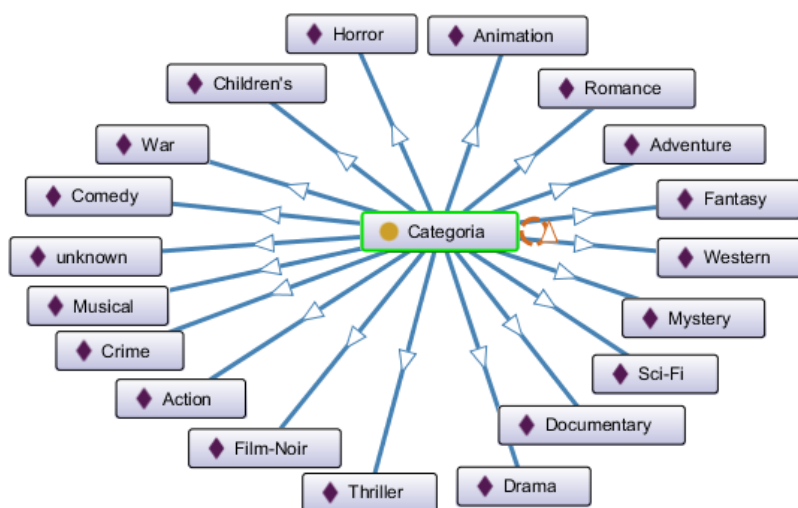
Após as modificações apresentadas o conjunto de dados foi carregado para a ontologia através de um *parser* desenvolvido em linguagem de programação Java®. O código desenvolvido é capaz de inserir de forma automática as instâncias, bem como as propriedades de dado e objeto. Para isso utilizou-se a biblioteca Apache Jena®⁴ que facilita a integração entre o Java e ontologias no formato OWL.

⁴ <https://jena.apache.org/>

5.3 VISUALIZANDO INSTÂNCIAS

As primeiras instâncias criadas foram da classe **Categoria**. Apesar de o modelo suportar uma estrutura de categorias e subcategorias, no conjunto de dados não há subcategorias. Desse modo, criou-se um conjunto de categorias de mesmo nível. A Figura 29 demonstra as 19 categorias instanciadas a partir dos dados.

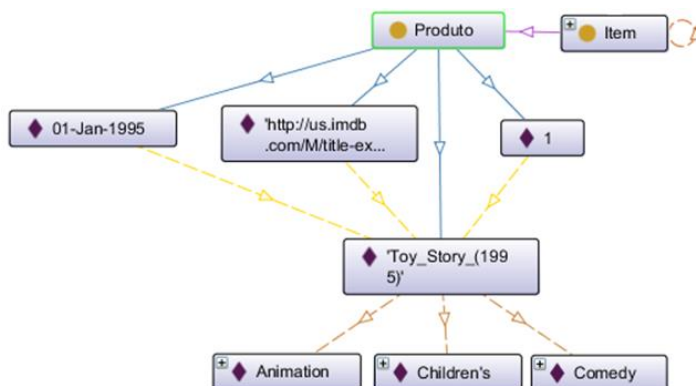
Figura 29 - Instâncias da classe **Categoria**.



Fonte: Autor.

A segunda classe a ser instanciada foi a **Item**. Como há uma grande quantidade de itens a Figura 30 demonstra apenas o primeiro item do conjunto de dados, com intuito de facilitar a visualização. Também é possível perceber a relação do item com as categorias abordadas anteriormente. Fica evidente ainda a estrutura de composição de itens, onde o item “Toy Story” é composto por uma data de lançamento, seu endereço no site de crítica IMDB, e o seu identificador.

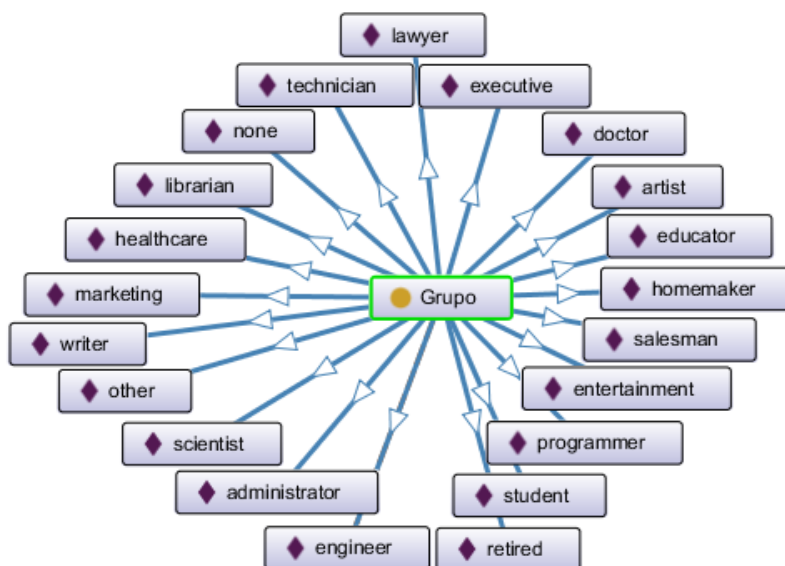
Figura 30 - Exemplo de instância da classe **Item**.



Fonte: Autor.

As instâncias da classe grupo podem ser visualizadas na Figura 31. Os 21 grupos instanciados referem-se a profissão dos usuários obtidas a partir do conjunto de dados considerando o campo “occupation”.

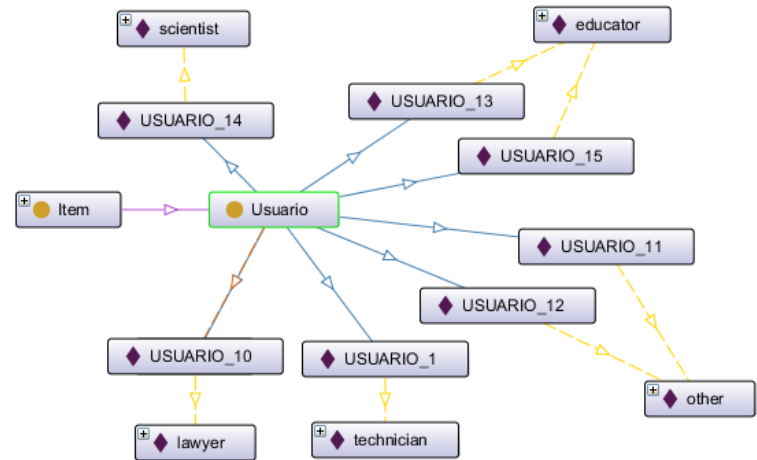
Figura 31 - Instâncias da classe **Grupo**.



Fonte: Autor.

Com as instâncias da classe **Grupo** inseridas, é possível demonstrar o relacionamento em usuários e seus grupos. Assim como na demonstração dos itens, a Figura 32 limita o número de usuário com o objetivo de facilitar a visualização.

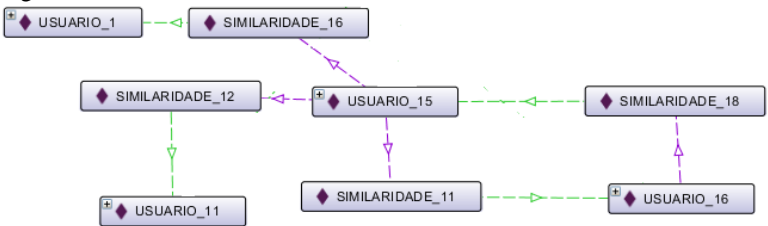
Figura 32 - Relacionamentos entre usuários e grupos.



Fonte: Autor.

Após a instanciação dos usuários é possível calcular e adicionar instâncias à classe **Similaridade**. Esta classe é responsável por determinar a quão semelhantes são os usuários e, deste modo, permitir recomendações baseados no perfil. A Figura 33 apresenta duas instâncias de usuários e três similaridades. Vale lembrar que o valor da similaridade entre usuários é armazenado na propriedade de *valor_similaridade*.

Figura 33 - Relacionamentos entre usuários através da similaridade.



Fonte: Autor.

Com isso concluem-se as sete etapas propostas pela Metodologia 101. Através da definição das classes, propriedades e restrições de propriedades é possível representar o domínio de um sistema de recomendação. As instâncias criadas foram originadas do conjunto de dados fornecido pelo site *MovieLens*®, porém, o modelo é capaz de representar qualquer outra base de forma genérica.

Vale salientar que para chegar a este modelo cada etapa foi realizada diversas vezes, de forma iterativa, fazendo aprimoramentos a cada ciclo.

5.4 RESPONDENDO AS QUESTÕES DE COMPETÊNCIA

A presente sessão demonstra alguns exemplos de consultas que a base de conhecimento suporta. As consultas demonstradas a seguir foram definidas na sessão 4.1, que determina o escopo da ontologia indicando quais perguntas a mesma deveria ser capaz de responder.

As consultas serão apresentadas em duas partes, sendo a primeira em forma de quadros que contém uma consulta em formato SPARQL⁵. Na segunda parte o retorno da consulta é apresentado em forma de tabela, omitindo alguns elementos para torná-las menos extensas.

Além das consultas apresentadas, a ontologia permite a realização de outras perguntas que poderiam ser exploradas no domínio de Sistemas de Recomendação.

5.4.1 Qual o grupo (por exemplo, por profissão, por interesses, entre outros) que determinado usuário pertence?

Esta pode ser vista como uma pergunta simples e direta, mas comumente realizada quando se busca um melhor entendimento de uma situação, neste caso, a obtenção de um ou mais grupos a que determinado usuário esteja vinculado. A consulta apresentada no Quadro 22 produz o resultado detalhado na Tabela 12.

⁵ SPARQL é uma linguagem de consulta e um protocolo para acesso a arquivos em formato RDF e OWL elaborado pelo W3C.

Quadro 22 - Consulta em linguagem SPARQL relativa à pergunta 5.4.1.

```
SELECT ?Grupo
WHERE {
    o:USUARIO_1 o:fazParte ?grupo
}
```

Fonte: Autor.

Tabela 12- Resultado da consulta do Quadro 22.

Grupo
<i>Technician</i>

Fonte: Autor.

5.4.2 Quais as categorias preferidas pelo usuário?

Esta pergunta listar as categorias que o usuário tem mais apreço. Pode ser útil para gerar recomendações com base nas preferências do usuário. A consulta apresentada no Quadro 23 produz o resultado detalhado na Tabela 13.

Quadro 23 - Consulta em linguagem SPARQL relativa à pergunta 5.4.2.

```
SELECT ?categoria
WHERE {
    o:USUARIO_1 o:temPreferencia ?categoria
}
```

Fonte: Autor.

Tabela 13 - Resultado da consulta do Quadro 23.

Categoria
<i>Action</i>
<i>War</i>

Fonte: Autor.

5.4.3 Quais itens sofreram alguma interação por determinado usuário?

Questão fundamental para gerar recomendações, sendo parte da informação necessária para calcular a similaridade entre usuários. A consulta apresentada no Quadro 24 produz o resultado detalhado na Tabela 14.

Quadro 24 - Consulta em linguagem SPARQL relativa à pergunta 5.4.3.

```
SELECT ?item (str(?nota) as ?valor_avaliacao)
WHERE {
    o:USUARIO_1 o:usuarioInteracao ?av .
    ?av o:interacaoItem ?item .
    ?av o:valorAvaliacao ?nota
}
```

Fonte: Autor.

Tabela 14 - Resultado da consulta do Quadro 24.

Item	Valor de avaliação
Ace_Ventura:_Pet_Detective_(19994)	3
Aliens_(1986)	5
...	...
Wizard_of_Oz,_The_(1939)	4

Fonte: Autor.

5.4.4 Quais são os itens mais bem avaliados?

Questão útil para gerar recomendações mais genéricas, como a lista dos 10 itens mais bem avaliados. A consulta que resolve esta questão consta no Quadro 25, e o resultado pode ser observado na Tabela 15.

Quadro 25 - Consulta em linguagem SPARQL relativa à pergunta 5.4.4.

```
SELECT ?item (SUM(?valor_avaliacao)/COUNT(?item ) AS ?media)
WHERE {
    ?usuario o:usuarioInteracao ?interacao .
    ?interacao o:valorAvaliacao ?valor_avaliacao .
    ?interacao o:interacaoItem ?item
}
GROUP BY ?item
```

Fonte: Autor.

Tabela 15 - Resultado da consulta do Quadro 25.

Item	Média
Wizard_of_Oz,_The_(1939)	4.37
Antonia’s_Line_(1995)	4.25
...	...
Copycat_(1995)	2

Fonte: Autor.

5.4.5 Quais são os itens mais bem avaliados em determinada categoria?

Semelhante à questão anterior, porém permite fazer recomendações dos itens mais bem avaliados de determinada categoria, por exemplo, os 10 itens principais (melhor avaliação) da categoria comédia. O Quadro 26 apresenta a consulta e a Tabela 16 o resultado obtido.

Quadro 26 - Consulta em linguagem SPARQL relativa à pergunta 5.4.5.

SELECT ?item (SUM(?valor_avaliacao)/COUNT(?item) AS ?media) WHERE { ?usuario o:usuarioInteracao ?interacao . ?interacao o:valor_avaliacao_numerica ?valor_avaliacao. ?interacao o:interacaoItem ?item . ?item o:temCategoria o:Comedy } GROUP BY ?item
--

Fonte: Autor.

Tabela 16 - Resultado da consulta do Quadro 26.

Item	Média
Get_Shorty_(1995)	4.37
Babe_(1995)	4
...	...
From_Dusk_Till_Dawn_(1996)	2.6

Fonte: Autor.

5.4.6 Quais são os itens que sofreram mais interações?

Questão que evidencia os itens mais populares. Explicita o comportamento dos usuários, tornando uma informação útil para a tomada de decisão. A consulta apresentada no Quadro 27 produz o resultado detalhado na Tabela 17.

Quadro 27- Consulta em linguagem SPARQL relativa à pergunta 5.4.6.

```
SELECT ?item (count( ?interacao) as ?numero_avaliacoes)
WHERE {
    ?interacao o:interacaoItem ?item
}
GROUP BY ?item
ORDER BY DESC (?numero_avaliacoes)
```

Fonte: Autor.

Tabela 17 - Resultado da consulta do Quadro 27.

Item	Número de avaliações
Toy_Story_(1995)	11
Twelve_Monkeys_(1995)	11
...	...
Four_Rooms_(1995)	1

Fonte: Autor.

5.4.7 Quais são os itens que sofreram mais interações em determinada categoria?

Semelhante à questão anterior, porém define que categorias são mais populares. É também uma questão importante para mapear o comportamento dos usuários. A consulta apresentada no Quadro 28 produz o resultado detalhado na Tabela 18.

Quadro 28 - Consulta em linguagem SPARQL relativa à pergunta 5.4.7.

```
SELECT ?item (count( ?interacao) as ?numero_avaliacoes)
WHERE {
    ?interacao o:interacaoItem ?item .
    ?item o:temCategoria o:Comedy
}
GROUP BY ?item
ORDER BY DESC (?numero_avaliacoes)
```

Fonte: Autor.

Tabela 18 - Resultado da consulta do Quadro 28.

Item	Número de avaliações
Toy_Story_(1995)	11
Might_Aphrodite_(1995)	9
...	...
From_Dusk_Till_Dawn_(1996)	3

Fonte: Autor.

5.4.8 Quais são os perfis de usuários mais semelhantes a determinado usuário considerando seu perfil (USUARIO_1)?

Questão importante no processo de recomendação, pois define os usuários mais semelhantes a um usuário específico. A consulta que consta no Quadro 29 gera os resultados apresentados na Tabela 19.

Quadro 29 - Consulta em linguagem SPARQL relativa à pergunta 5.4.10.

```
SELECT ?usuario ?valor
WHERE
{
    o:USUARIO_1 ?usuarioSimilaridade ?similaridade .
    ?similaridade o:valor_similaridade ?valor .
    ?similaridade o:similaridadeUsuario ?usuario
}
ORDER BY DESC (?valor)
```

Fonte: Autor.

Tabela 19 - Resultado da consulta do Quadro 29.

Usuário	Valor
USUARIO_15	0.44
USUARIO_18	0.42
...	...
USUARIO_7	-0.38

Fonte: Autor.

5.5 GERANDO RECOMENDAÇÕES

Esta etapa demonstra os resultados das recomendações geradas a partir da interação entre as ontologias de tarefa e de domínio. Mais especificamente, as recomendações são obtidas a partir da execução dos métodos da ontologia de tarefa. A ontologia de tarefa utiliza como subsídio para gerar as recomendações o conhecimento contido na ontologia de domínio, ou seja, usuários, interações e itens.

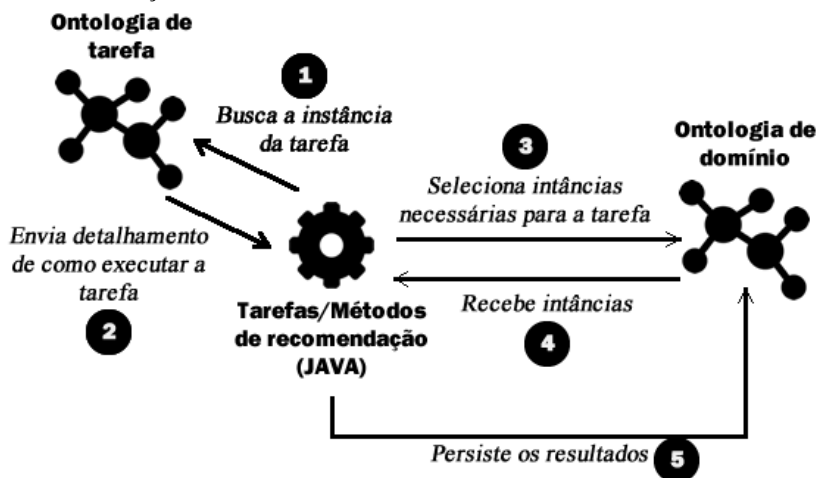
A utilização das duas ontologias é necessária, pois apesar da ontologia de domínio ser capaz de responder as questões de competência definidas, a mesma isoladamente não tem o conhecimento de como executar as etapas para que recomendações sejam efetuadas. Assim, configura-se uma estrutura onde é armazenado o conhecimento do domínio de recomendações (ontologia de domínio) e outra que detém o conhecimento de como gerar as recomendações (ontologia de tarefa).

A Figura 34 apresenta o fluxo e as interações entre a ontologia de tarefa, domínio e as tarefas implementadas.

Para esta etapa foi desenvolvida uma aplicação em Java® (nível de protótipo) responsável por conectar as ontologias e propiciar recomendações.

No primeiro passo a aplicação efetua uma requisição de uma tarefa através de uma consulta em linguagem SPARQL. A resposta da requisição, no passo dois contém a estrutura de uma tarefa, seus métodos e informações específicas de como a mesma deve ser executada.

Figura 34 - Fluxo, interações e componentes envolvidos para o processamento de recomendações.



Fonte: Autor.

Após a obtenção da tarefa, no passo 3 é realizada uma pesquisa na ontologia de domínio de modo que todas as informações necessárias para a recomendação sejam obtidas. Novamente, a interação é realizada através da linguagem SPARQL. O passo 4 representa o retorno das instâncias solicitadas no passo anterior e que irão propiciar a execução da tarefa. Por fim, com o conhecimento sobre a tarefa a ser executada e os indivíduos, a aplicação executa os métodos e persiste os resultados na ontologia de domínio conforme passo 5.

Nas próximas seções serão detalhadas a execução dos métodos contidos na ontologia de tarefa, assim como, os resultados que permitem prover recomendações para as abordagens desenvolvidas. Para executar de forma genérica os métodos de uma tarefa desenvolveu-se uma aplicação utilizando a linguagem de programação Java® através do conceito de reflexão. Esta aplicação recebe o nome de uma tarefa e a partir disso busca na ontologia de tarefa os métodos e os demais elementos de modo que esta possa ser resolvida. A Figura 35 apresenta o fluxo de execução da aplicação.

Figura 35 - Estrutura genérica para execução de tarefas.



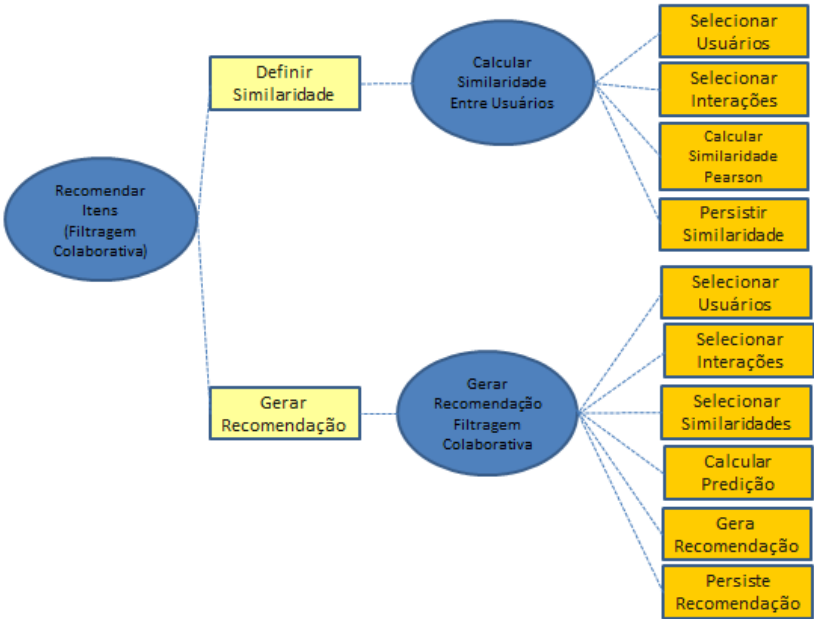
Fonte: Autor

Como mencionado anteriormente as abordagens demográfica, baseada em conteúdo, baseada em comunidade, filtragem colaborativa e híbrida foram desenvolvidas para a avaliação da integração entre a ontologia de tarefa e de domínio. Para fins de demonstração, serão discutidas nas próximas seções as abordagens de filtragem colaborativa e demográfica.

5.5.1 Recomendando através da Filtragem Colaborativa

A filtragem colaborativa é a abordagem mais utilizada nos sistemas de recomendação. O funcionamento da abordagem é relativamente simples, sendo necessário primeiramente calcular a similaridade entre os usuários. A partir da similaridade entre os mesmos é possível calcular a predição, que visa prever a nota de avaliação que o usuário dará para determinado item. Quanto maior o valor de predição, maiores são as chances de o usuário apreciar o item recomendado. A Figura 37 apresenta a estrutura da tarefa e dos métodos executados na filtragem colaborativa.

Figura 36 - Estrutura da tarefa “Recomendar Itens” (Filtragem Colaborativa).



Fonte: Autor.

É possível perceber que existem métodos primitivos que se repetem na estrutura apresentada. As tarefas, representadas por elipses azuis, também pode compor outras tarefas mais complexas. Está é uma característica importante e faz com que um método ou tarefa sejam reutilizados quantas vezes for necessário.

O Quadro 30 faz uma descrição dos métodos que compõem a tarefa “Recomendar Itens”.

Quadro 30 - Detalhamento da tarefa apresentada na Figura 36.

Método Complexo	Tarefa	Método Primitivo
Definir Similaridade	Calcular similaridade entre usuários	Método: selecionarUsuario. Saída: Lista de usuários. Descrição: Seleciona todos os usuários da ontologia.
		Método: selecionarInteracoes. Entrada: Lista de usuários. Saída: Lista de interações. Descrição: Seleciona todas as interações (numéricas) dos usuários.
		Método: calcularSimilaridade. Entrada: Lista de usuários, lista de interações. Saída: Lista de similaridades.
		Método: persisteSimilaridade. Entrada: Lista de similaridades. Descrição: Persiste as similaridades na ontologia.
Gerar Recomendação	Gerar recomendação (filtragem colaborativa)	Método: selecionarUsuário. Saída: Lista de usuários. Descrição: Seleciona todos os usuários da ontologia.
		Método: selecionarInteracoes. Entrada: Lista de usuários. Saída: Lista de interações. Descrição: Seleciona todas as interações (numéricas) dos usuários.
		Método: selecionarSimilaridades. Entrada: Lista de usuários. Saída: Lista de similaridades. Descrição: Seleciona todas as interações (numéricas) dos usuários.
		Método: calcularPredicao. Entrada: Lista de usuários, lista de interações, lista de

		similaridades. Saída: Lista de predições. Descrição: calcula a predição, base para as recomendações geradas.
		Método: geraRecomendacao. Entrada: Lista de usuários, lista de predições. Saída: Lista de recomendações. Descrição: utiliza as predições para gerar recomendações.
		Método: persisteRecomendacao. Entrada: Lista de recomendações. Descrição: Fim da tarefa, as recomendações são armazenadas na ontologia.

Fonte: Autor.

A primeira tarefa calcula a similaridade dos pares de usuários, informação fundamental para as etapas posteriores. No exemplo a seguir a equação de similaridade utilizada foi a correlação de Pearson, cujo coeficiente pode variar de -1 para usuários totalmente opostos até 1 para usuários com preferências idênticas.

A Figura 37 e Figura 38 apresentam os pares de usuários com os maiores e menores coeficientes de similaridade, respectivamente.

Figura 37 - Usuários mais similares.

usuario1	valor_sim	
USUARIO_10	"0.25847509057750084"	USUARIO_1
USUARIO_13	"0.25375819212199435"	USUARIO_1
USUARIO_10	"0.24796587492955036"	USUARIO_16
USUARIO_11	"0.23952989809191097"	USUARIO_1
USUARIO_1	"0.23030376370584307"	USUARIO_14
USUARIO_16	"0.21724221067280566"	USUARIO_12

Fonte: Autor.

Figura 38 - Usuários menos similares.

usuario1	valor_sim	
USUARIO_12	"0.043992199152109994"	USUARIO_11
USUARIO_16	"0.05141089195719047"	USUARIO_15
USUARIO_10	"0.08248526088099614"	USUARIO_15
USUARIO_11	"0.09078238917872587"	USUARIO_10
USUARIO_11	"0.09323294021064707"	USUARIO_14
USUARIO_15	"0.13025299731797296"	USUARIO_12

Fonte: Autor.

As figuras anteriores demonstram os maiores e menores coeficientes dentre todos os usuários da base de conhecimento. Porém também é possível determinar estes valores considerando um usuário específico, como pode ser observado na Figura 39 e Figura 40.

Figura 39 - Usuários menos similares ao USUARIO_1.

usuario_x	similaridade	usuario_y
USUARIO_1	"0.01917019087231611"	USUARIO_29
USUARIO_1	"0.040978915581998746"	USUARIO_3
USUARIO_1	"0.07083174431201295"	USUARIO_30
USUARIO_1	"0.07395776612015885"	USUARIO_4
USUARIO_1	"0.07675072900397224"	USUARIO_19
USUARIO_1	"0.09654943751356551"	USUARIO_27
USUARIO_1	"0.09852547159666272"	USUARIO_9
USUARIO_1	"0.11237173813431243"	USUARIO_2
USUARIO_1	"0.12321234291274229"	USUARIO_20
USUARIO_1	"0.13923479683237364"	USUARIO_12
USUARIO_1	"0.1466370960258472"	USUARIO_28
USUARIO_1	"0.159990489752878"	USUARIO_24
USUARIO_1	"0.16043049968631934"	USUARIO_16

Fonte: Autor.

Figura 40 - Usuários mais similares ao USUARIO_1.

usuario_x	similaridade	usuario_y
USUARIO_1	"0.38758003788646495"	USUARIO_18
USUARIO_1	"0.2945791671219222"	USUARIO_6
USUARIO_1	"0.28326098757864504"	USUARIO_23
USUARIO_1	"0.2691746004221829"	USUARIO_7
USUARIO_1	"0.25847509057750084"	USUARIO_10
USUARIO_1	"0.25375819212199435"	USUARIO_13
USUARIO_1	"0.23952989809191097"	USUARIO_11
USUARIO_1	"0.23030376370584327"	USUARIO_14
USUARIO_1	"0.22665026492408868"	USUARIO_26
USUARIO_1	"0.22345690643603366"	USUARIO_5
USUARIO_1	"0.2156234115632919"	USUARIO_22
USUARIO_1	"0.2097914889665735"	USUARIO_21
USUARIO_1	"0.2072815916106533"	USUARIO_25

Fonte: Autor.

O próximo passo, calcular a predição, utiliza dados dos usuários mais similares. Há dois modos de efetuar o corte dos usuários mais similares. O passo primeiro consiste em determinar um valor mínimo para o coeficiente de similaridade. Neste caso, é possível que nenhum usuário possua um coeficiente maior do que o estipulado, tornando assim a recomendação inviável. O segundo modo visa determinar um número de usuários mais similares, mas também pode haver problemas, caso os maiores coeficientes possuam um valor baixo. No método desenvolvido optou-se por considerar os cinco usuários mais similares para realizar os cálculos de predição. A Figura 41 apresenta os cinco usuários mais similares ao USUARIO_1 com os respectivos valores de similaridade.

Figura 41 - Os cinco usuários mais similares ao USUARIO_1.

usuario_x	similaridade	usuario_y
USUARIO_1	"0.38758003788646495"	USUARIO_18
USUARIO_1	"0.2945791671219222"	USUARIO_6
USUARIO_1	"0.28326098757864504"	USUARIO_23
USUARIO_1	"0.2691746004221829"	USUARIO_7
USUARIO_1	"0.25847509057750084"	USUARIO_10

Fonte: Autor.

A partir da seleção do conjunto reduzido dos usuários mais similares, e com as informações de suas interações contidas na base é possível aplicar o cálculo de predição. A Figura 42 demonstra uma lista de itens seguidos pelo valor de predição.

Figura 42 - Lista de itens com seus valores de predição.

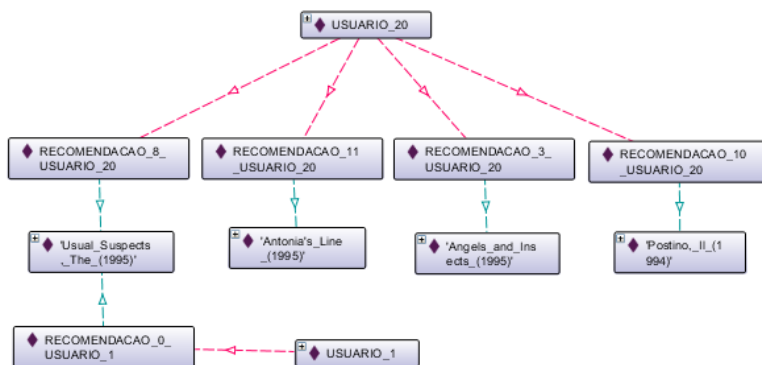
item	valor_predicao
'Return_of_the_Jedi_(1983)'	"4.714102999143887"
'2001:_A_Space_Odyssey_(1968)'	"4.696262217797676"
'Big_Night_(1996)'	"4.679634231431713"
'Shawshank_Redemption,_The_(1994)'	"4.639663782726507"
'Close_Shave,_A_(1995)'	"4.635486748817207"
'GoodFellas_(1990)'	"4.634885198924328"
'Great_Escape,_The_(1963)'	"4.514740334245616"
'Alice_in_Wonderland_(1951)'	"4.514740334245616"
'Young_Frankenstein_(1974)'	"4.514421890891298"
'To_Kill_a_Mockingbird_(1962)'	"4.490206478858204"
'Usual_Suspects,_The_(1995)'	"4.475965414084235"
'Pinocchio_(1940)'	"4.472530966702998"
'Night_on_Earth_(1991)'	"4.472530966702998"

Fonte: Autor.

O valor de predição define que item será recomendado. Porém, vale salientar que nem toda recomendação é baseada no valor de predição, como por exemplo, a recomendação de lançamentos que é norteadada apenas pela data de lançamento. O trabalho desenvolvido gera recomendações de todos os itens que tiveram seu valor de predição calculado, deixando a cargo da aplicação que acessar a base de conhecimento, decidir qual o valor mínimo admitido.

Por fim, as recomendações são geradas e persistidas na ontologia de domínio. A Figura 43 apresenta um exemplo contido na ontologia de domínio que contém as recomendações geradas, os itens que a compõem e os usuários a quem são destinadas.

Figura 43 - Visão geral das recomendações na ontologia de domínio.



Fonte: Autor.

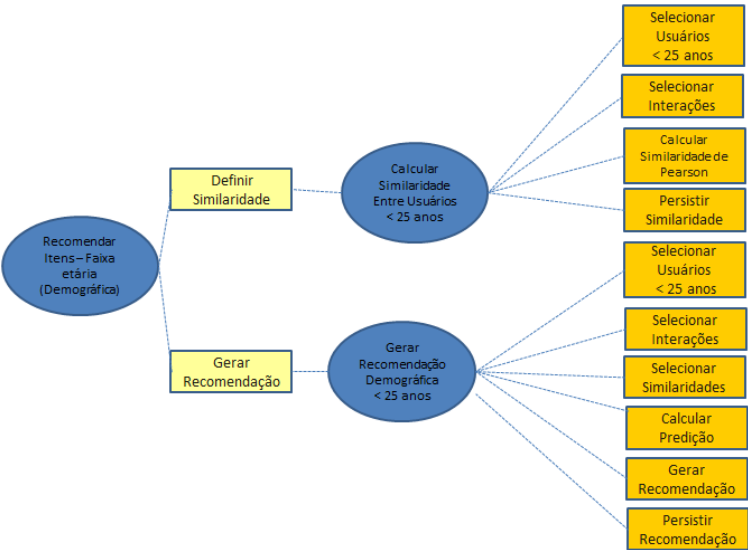
Após a conclusão da última etapa da tarefa de recomendação, pode-se afirmar que o modelo se comportou de forma adequada. A ontologia de tarefa é capaz de executar os passos necessários para gerar as recomendações de maneira correta. E evidenciou uma característica da filtragem colaborativa, o rápido crescimento da base, que armazena interações e as similaridades entre os usuários.

5.5.2 Recomendando através da abordagem demográfica

O processo de recomendação da abordagem demográfica é semelhante ao apresentado na filtragem colaborativa, porém este considera também dados demográficos para realizar recomendações. A

Figura 44 apresenta a estrutura de tarefa da de recomendação demográfica e seus métodos.

Figura 44 - Estrutura da tarefa “Recomendar Itens” (Demográfica).



Fonte: Autor.

A estrutura de tarefa apresentada gera recomendações para usuários com menos de 25 anos. Apesar do método primitivo declarar a idade do usuário, a implementação poderia ser generalizada para permitir qualquer idade, ou mesmo, outros comportamentos, tais como, maior idade ou um intervalo de idades. A descrição dos métodos que compõem a tarefa pode ser observada no Quadro 31.

Quadro 31 - Detalhamento da tarefa apresentada na Figura 44.

.Método Complexo	Tarefa	Método Primitivo
Definir Similaridade	Calcular Similaridade entre usuários menores de 25 anos	Método: selecionarUsuárioMenor25. Saída: Lista de usuários. Descrição: Seleciona os usuários com menos de 25 anos.
		Método: selecionarInteracoes. Entrada: Lista de usuários. Saída: Lista de interações. Descrição: Seleciona todas as

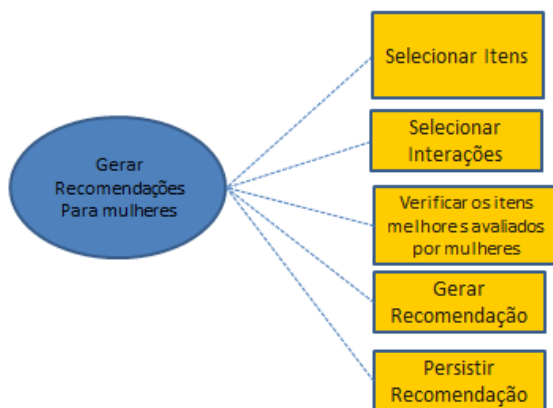
		interações (numéricas) dos usuários.
		Método: calcularSimilaridade. Entrada: Lista de usuários, lista de interações. Saída: Lista de similaridades.
		Método: persisteSimilaridade. Entrada: Lista de similaridades. Descrição: Persiste as similaridades na ontologia.
Gerar Recomendação	Gerar Recomendação filtragem colaborativa para usuários com menos de 25 anos	Método: selecionarUsuárioMenor25. Saída: Lista de usuários. Descrição: Seleciona os usuários com menos de 25 anos.
		Método: selecionarInteracoes. Entrada: Lista de usuários. Saída: Lista de interações. Descrição: Seleciona todas as interações (numéricas) dos usuários.
		Método: selecionarSimilaridades. Entrada: Lista de usuários. Saída: Lista de similaridades. Descrição: Seleciona todas as interações (numéricas) dos usuários.
		Método: calcularPredicao. Entrada: Lista de usuários, lista de interações, lista de similaridades. Saída: Lista de predições. Descrição: Calcula a predição, base para as recomendações geradas.
		Método: geraRecomendacao. Entrada: Lista de usuários, lista de predições. Saída: Lista de recomendações. Descrição: Utiliza as predições para gerar recomendações.
		Método: persisteRecomendacao. Entrada: Lista de recomendações. Descrição: Fim da tarefa, as recomendações são armazenadas na ontologia.

Fonte: Autor.

A tarefa apresentada segue os mesmos passos da filtragem colaborativa, que já foi detalhada na sessão anterior. Porém, existe ainda a possibilidade de gerar recomendações sem determinar a similaridade e a predição como na tarefa anterior.

Uma recomendação pode ser baseada apenas em dados demográficos, deixando de lado a análise do perfil do usuário. Deste modo, é possível definir algumas recomendações para pessoas de determinado sexo, idade, local, entre outros. A Figura 45 exemplifica uma entre varias tarefas de recomendação utilizando apenas os dados demográficos.

Figura 45 - Estrutura da tarefa “Recomendar para Mulheres”.



Fonte: Autor.

Seguindo essa linha de raciocínio a Tabela 20, a Tabela 21 e a Tabela 22 apresentam os dez itens mais bem avaliados, em que a primeira tabela não faz distinção do sexo dos avaliadores, enquanto as outras duas tabelas são compostas por avaliações de mulheres e homens, respectivamente.

Tabela 20 - Os 10 itens mais bem avaliados (Geral).

Posição	Item	Média Avaliações
1	Close Shave, A (1995)	4.4911
2	Schindler's List (1993)	4.4664
3	Wrong Trousers, The (1993)	4.4661
4	Casablanca (1942)	4.4568
5	Wallace & Gromit...	4.4478
6	Shawshank Redemption, The (1994)	4.4452
7	Rear Window (1954)	4.3876
8	Usual Suspects, The (1995)	4.3858
9	Star Wars (1977)	4.3585
10	12 Angry Men (1957)	4.3440

Fonte: Autor.

Tabela 21 - Os 10 itens mais bem avaliados (Mulheres)

Posição	Item	Média Avaliações
1	Schindler's List (1993)	4.6329
2	Close Shave, A (1995)	4.6316
3	Shawshank Redemption, The (1994)	4.5625
4	Wallace & Gromit...	4.5333
5	Shall We Dance? (1996)	4.5294
6	Friday (1995)	4.5000
7	Wrong Trousers, The (1993)	4.4444
8	Some Folks Call It a Sling Blade (1993)	4.4167
9	Sling Blade (1996)	4.4118
10	Secrets & Lies (1996)	4.4074

Fonte: Autor.

Tabela 22 - Os 10 itens mais bem avaliados (Homens).

Posição	Item	Média Avaliações
1	Casablanca (1942)	4.4734
2	Wrong Trousers, The (1993)	4.4725
3	Close Shave, A (1995)	4.4624
4	Third Man, The (1949)	4.4333
5	Wallace & Gromit...	4.4231
6	Paths of Glory (1957)	4.4194
7	Shawshank Redemption, The (1994)	4.4110
8	Schindler's List (1993)	4.4064
9	Rear Window (1954)	4.4000
10	Usual Suspects, The (1995)	4.3991

Fonte: Autor.

As tabelas apresentadas demonstram uma diferenciação na avaliação dos itens quando divididos por gêneros. Dos dez itens mais bem avaliados de forma geral, oito constavam na lista das preferências masculinas. Já na tabela da amostra feminina existiam apenas cinco dos dez itens mais bem avaliados de forma geral. Outro aspecto importante é que mesmo possuindo algumas semelhanças a posição dos filmes variou bastante de uma tabela para outra.

Além de gênero o conjunto de dados disponibiliza a idade, outro atributo demográfico importante no para o processo de recomendação. As Tabela 23, Tabela 24 e Tabela 25 demonstram a diferença de preferencias dos usuários divididos por faixas etárias.

Tabela 23 - Os 10 itens mais bem avaliados (Menores de 25 anos).

Posição	Item	Média Avaliações
1	Paradise Lost: The Child Murders at Robin Hood Hil...	5.0000
2	Beautiful Thing (1996)	5.0000
3	Top Hat (1935)	5.0000
4	Wallace & Gromit...	4.7500
5	Affair to Remember, An (1957)	4.6667
6	Shawshank Redemption, The (1994)	4.6400
7	Casablanca (1942)	4.6216
8	Titanic (1997)	4.5313
9	Waiting for Guffman (1996)	4.5000
10	Dr. Strangelove or: How I Learned to Stop Worrying...	4.5000

Fonte: Autor.

Tabela 24 - Os 10 itens mais bem avaliados (De 25 a 50 anos).

Posição	Item	Média Avaliações
1	Close Shave, A (1995)	4.5000
2	Wallace & Gromit...	4.5000
3	Wrong Trousers, The (1993)	4.4861
4	Schindler's List (1993)	4.4830
5	Casablanca (1942)	4.4465
6	Usual Suspects, The (1995)	4.4276
7	Shawshank Redemption, The (1994)	4.4214
8	Rear Window (1954)	4.3684
9	Third Man, The (1949)	4.3571
10	Paths of Glory (1957)	4.3571

Fonte: Autor.

Tabela 25 - Os 10 itens mais bem avaliados (Maiores de 50 anos).

Posição	Item	Média Avaliações
1	Cemetery Man (Dellamorte Dellamore) (1994)	5.0000
2	Half Baked (1998)	5.0000
3	Pillow Book, The (1995)	5.0000
4	Shall We Dance? (1996)	4.8571
5	Wrong Trousers, The (1993)	4.7143
6	Priest (1994)	4.6667
7	Vanya on 42nd Street (1994)	4.6667
8	Indian Summer (1996)	4.6667
9	Kull the Conqueror (1997)	4.6000
10	Schindler's List (1993)	4.5806

Fonte: Autor.

É perceptível a diferença dos itens contidos nas tabelas acima. Contando os trinta itens presentes nas três tabelas apenas 5 se repetem. Este tipo de comportamento promove espaço para a utilização de recomendações demográficas com intuito de atender conjuntos de usuários com preferências distintas.

Também é possível utilizar mais de um atributo demográfico para selecionar um grupo específico. Este refinamento deve ser realizado com atenção, pois tende a produzir tabelas mais esparsas e muitas vezes, pode gerar recomendações pouco relevantes, uma vez que se baseia em um número pequeno de avaliações. A Tabela 26 demonstra este fenômeno, onde todos os dez itens possuem nota média máxima (5), porém, estes itens sofreram poucas avaliações.

Tabela 26 - Os 10 itens mais bem avaliados (Mulheres com mais de 50 anos).

Posição	Item	Média Avaliações	Quant. De avaliações
1	Shall We Dance? (1996)	5.0000	4
2	Firm, The (1993)	5.0000	1
3	Searching for Bobby Fischer (1993)	5.0000	1
4	Client, The (1994)	5.0000	1
5	Man Without a Face, The (1993)	5.0000	1
6	E.T. the Extra-Terrestrial (1982)	5.0000	2
7	Fried Green Tomatoes (1991)	5.0000	4
8	Room with a View, A (1986)	5.0000	1
9	Peacemaker, The (1997)	5.0000	2
10	MatchMaker, The (1997)	5.0000	1

Fonte: Autor.

5.6 CONSIDERAÇÕES FINAIS

O capítulo 5 trouxe detalhes sobre a aplicação do modelo bem como as discussões sobre os resultados obtidos.

Para ser aplicado o modelo proposto foi populado através de parte do *dataset* MovieLens 100k que contém informações sobre filmes, usuário e suas avaliações. Sendo assim os dados contidos na base foram adicionados a ontologia de domínio.

Já contendo as informações foram realizadas consultas via *SPARQL* com intuito de verificar se a ontologia de domínio era capaz de responder as perguntas de competência, característica fundamental do modelo.

Por fim a ontologia de tarefa foi testada através da geração de recomendações com base na ontologia de domínio. As tarefas de recomendação foram implementadas na linguagem de programação Java®. Os testes foram realizados em duas abordagens, filtragem colaborativa e demográfica, e para realização sua realização também foi necessário definir as tarefas a serem realizadas em ambas as abordagens, expondo assim suas estruturas e características. Através dos testes foi possível comprovar a eficácia do modelo proposto em realizar as atividades previstas.

6 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

6.1 CONSIDERAÇÕES FINAIS

Os Sistemas de Recomendação vem ganhando destaque atualmente devido principalmente a sua capacidade de auxiliar em tarefas que lidem com grandes quantidades de informação, ou tenham características muito complexas.

Como consequência a esta grande visibilidade, os Sistemas de Recomendação vem evoluindo e se adaptando para adequarem-se a uma variedade de cenários. Uma solução relativamente nova é a integração entre estes sistemas e estruturas capazes de representar conhecimento.

Desta maneira, os Sistemas de Recomendação baseados em conhecimento vem com a proposta de utilizar conhecimento de domínio para gerar recomendações. Porém, estas estruturas de representação de conhecimento não tem a capacidade de suportar as tarefas necessárias contidas no processo de recomendação.

O objetivo geral consistiu na proposição de um modelo de representação do conhecimento voltado à tarefa intensiva em conhecimento de recomendação.

De modo a cumprir o objetivo geral este trabalho se concentrou em alguns objetivos específicos iniciando pela revisão da literatura para subsidiar a proposição e desenvolvimento do modelo. Após a revisão da literatura percebeu-se que ontologias, mais especificamente de domínio e de tarefas, poderiam atender esta demanda. Sendo assim, este trabalho propõe um modelo composto de uma ontologia de domínio e de tarefa que se integram para gerar recomendações e auxiliar no processo de tomada de decisão.

Para chegar ao modelo proposto primeiramente a ontologia de domínio foi projetada através da aplicação da metodologia 101 considerando o conhecimento obtido na elaboração da revisão sobre os Sistemas de Recomendação. A ontologia de domínio foi construída de maneira que seja possível recomendar qualquer tipo de item, atuando de forma genérica. Um dos pontos fundamentais na metodologia é a definição de um conjunto de perguntas de competência, que possui a finalidade de avaliação a capacidade da ontologia em responder tais demandas. Vale ressaltar, que durante a pesquisa não foram encontradas ontologias voltadas a representação do domínio de Sistemas de Recomendação.

A ontologia de tarefa posteriormente desenvolvida possui o conhecimento de como executar as tarefas intensivas em conhecimento de recomendação. A ontologia compreende os passos que devem ser executados, as classes e métodos envolvidos, e os parâmetros que serão utilizados. Neste sentido, as diversas abordagens presentes na literatura de Sistemas de Recomendação, podem ser configuradas como tarefas.

Para demonstrar a viabilidade do modelo a ontologia de domínio foi populada com um conjunto de dados utilizados em um sistema de recomendação de filmes, provida pelo site *MovieLens*®. Com as instâncias na ontologia, as perguntas de competência podem ser avaliadas. No contexto deste trabalho considera-se que a ontologia de domínio uma estrutura adequada uma vez que o resultado foi positivo em todas as situações, considerando as perguntas de competência realizadas e o contexto em que foi aplicada.

Adicionalmente, uma ontologia de tarefa foi utilizada para gerar recomendações, através das abordagens de filtragem colaborativa e demográfica. Em ambas as situações a utilização da ontologia de tarefa conseguiu gerar recomendações a partir da base de conhecimento suportada pela ontologia de domínio.

De modo geral, pode-se considerar que o modelo proposto foi capaz de atender o objetivo geral e responder a pergunta de pesquisa deste trabalho. A ontologia de domínio se mostra consistente e adequada a representação da área de Sistemas de Recomendação, uma vez que as perguntas elaboradas na fase de especificação são passíveis de resposta. A ontologia de tarefa segue o conceito de divisão em tarefas e subtarefas em que cada uma destas é composta por um conjunto de métodos.

Neste sentido, a proposição da ontologia foi capaz de prover os subsídios para a conexão com a ontologia de domínio e desta maneira responder as tarefas de recomendação implementadas. As tarefas se mostram consistentes e são capazes de auxiliar no processo de tomada de decisão de usuários que de alguma maneira interajam com determinado sistema através de recomendações.

6.2 TRABALHOS FUTUROS

Durante o desenvolvimento deste trabalho algumas possibilidades de continuidade foram identificadas. O modelo proposto traz uma contribuição no contexto da representação semântica de Sistemas de Recomendação através da elaboração de uma ontologia de domínio. Além disso, a utilização de ontologias de tarefa constitui-se em um recurso adequado visando à integração entre as especificidades de determinada tarefa e o conhecimento de domínio.

No estado atual, o modelo fornece subsídios em como aprimorar e até mesmo abordar questões evidenciadas no decorrer deste trabalho que não foram consideradas.

Um primeiro ponto a considerar refere-se à avaliação da ontologia de domínio proposta em outros cenários visando identificar o grau de generalidade da mesma. A ontologia foi elaborada com foco no conceito de item, o que objetiva garantir que outros cenários voltados à recomendação possam ser atendidos. Todavia, considerando especificidades de outros domínios adaptações e/ou incrementos podem ser necessários.

Em um contexto mais computacional duas questões surgem e possibilitam a evolução do presente trabalho. A primeira refere-se ao aprimoramento dos métodos capazes de atender uma gama maior de tarefas de recomendação. A segunda envolve um estudo em relação a capacidade de raciocínio em ontologia de domínio e grandes bases de conhecimento, o que atualmente é um desafio. A tecnologia de Banco de Dados tem evoluído nos últimos anos, mas ainda demonstram limitações neste ponto. Estudos em como adequar soluções tecnológicas atuais para

suportar cenários complexos de representação de conhecimento aliados a capacidade de raciocínio demonstram potencial de pesquisa.

Por fim, vislumbra-se a proposição de um *framework* que facilite a integração de ontologias de domínio e de tarefa de maneira flexível. Para uma integração mais ampla o *framework* deveria considerar ainda ontologias de nível superior e ontologias de aplicação. Tal proposição tem como objetivo final facilitar o desenvolvimento de Sistemas Baseados em Conhecimento, possibilitando que especialistas de domínio participem na concepção como um todo desta classe de sistema.

REFERÊNCIAS

ADOMAVICIUS, Gediminas; TUZHILIN, Alexander. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. **Knowledge And Data Engineering**, S. I., v. 17, n. 6, p.734-749, jun. 2005.

AL-SHAMRI, Mohammad Yahya H.. Power coefficient as a similarity measure for memory-based collaborative recommender systems. **Expert Systems With Applications**, S. I., v. 41, n. 13, p.5680-5688, out. 2014.

Argollo, Rita Virginia et al. Web 2.0 como estruturante dos processos de produção e difusão científica em um grupo de pesquisa: o TWIKI e o GEC. **PERSPECTIVAS EM CIÊNCIA DA INFORMAÇÃO**, Abr 2010, vol.15, no.1, p.118-131.

MARTINS, Aline Freitas. **CONSTRUÇÃO DE ONTOLOGIAS DE TAREFA E SUA REUTILIZAÇÃO NA ENGENHARIA DE REQUISITOS**. 2009. 178 f. Dissertação (Mestrado) - Curso de Informática, Universidade Federal do Espírito Santo, Vitória, 2009.

AMATRIAIN, Xavier et al. Data Mining Methods for Recommender Systems. In: RICCI, Francesco et al. **Recommender Systems Handbook**. New York: Springer, 2011. Cap. 2. p. 39-71.

ALMEIDA, Mauricio B.; BAX, Marcello P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e construção. **Ciência da Informação**, Brasília, v. 32, n. 3, p. 7-20, set./dez. 2003.

ALMEIDA, M.B; SOUZA, R.R.; BAX, M.P. O grupo de pesquisa “Representação do Conhecimento, Ontologias e Linguagem”: atividades e perspectivas. **II Seminário de Pesquisa em Ontologia no Brasil**, IME - Rio de Janeiro, Set. / 2009.

BAMBINI, Riccardo; CREMONESI, Paolo; TURRIN, Roberto. A Recommender System for an IPTV Service Provider: a Real Large-Scale Production Environment. In: RICCI, Francesco et al. **Recommender Systems Handbook**. New York: Springer, 2011. Cap. 9. p. 299-331.

BOTHOS, Efthimios; APOSTOLOU, Dimitris; MENTZAS, Gregoris. Collective intelligence with web-based information aggregation markets: The role of market facilitation in idea management. **Expert Systems With Applications**, S. I., v. 39, n. 1, p.1333-1345, jan. 2012.

BOLLEN, Dirk et al. Understanding choice overload in recommender systems. In: **ACM CONFERENCE ON RECOMMENDER SYSTEMS, 10., 2010, Barcelona**. Proceedings of the fourth ACM conference on Recommender systems. New York: Acm, 2010. p. 63 - 70.

BREITMAN, Karin Koogan. **Web semântica: a internet do futuro**. Rio de Janeiro: LTC, 2006. xvii, 190 p.

BREITMAN, Karin Koogan; CASANOVA, Marco Antonio; RUSZKOWSKI, Walter. **Semantic Web: Concepts, Technologies and Applications**. London: Springer-Verlag, 2007. 327 p.

BURIANO, Luca et al. The Role of Ontologies in Context-Aware Recommender Systems. In: **INTERNATIONAL CONFERENCE ON MOBILE DATA MANAGEMENT, 7., 2006, Los Alamitos**. Proceedings of the 7th International Conference on Mobile Data Management. S. I.: Ieee, 2006. p. 80 - 82.

BURKE, Robin. Hybrid recommender systems: Survey and experiments. **User Modeling and User Adapted Interaction**, v.12, n. 6, p. 331–370, nov. 2002.

BURKE, Robin; FELFERNIG, Alexander; GÖKER, Mehmet H.. Recommender Systems: An Overview. **Ai Magazine**. S. I., p. 13-18. 2011.

BURKE, Robin; RAMEZANI, Maryam. Matching Recommendation Technologies and Domains. In: RICCI, Francesco et al. **Recommender Systems Handbook**. New York: Springer, 2011. p. 367-386.

CAMPOS, Maria Luiza de Almeida. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais **Ciência da Informação**, Brasília , v. 33, n. 1, abr. 2004 .

CARRER-NETO, Walter et al. Social knowledge-based recommender system: Application to the movies domain. **Expert Systems With Applications**, S. I., v. 39, n. 12, p.10990-11000, 15 set. 2012.

CASTILHO, L.V. ; LOPES, H. S. ; TACLA, C. A.. Modeling and Building an Ontology for Neuropediatric Physiotherapy Domain. In: **Hybrid Intelligent Systems**, 2008, Barcelona, Espanha. Proceedings of the 8th International Conference on Hybrid Intelligent Systems. Piscataway, EUA : IEEE Computer Press, v. 1, p. 210-215, 2008.

CAZELLA, S. C. NUNES, M. A. S. N. REATEGUI, E. B; A Ciência do Palpite: Estado da Arte em Sistemas de Recomendação. In: **JORNADA DE ATUALIZAÇÃO DE INFORMÁTICA**. Rio de Janeiro: PUC RIO, v. 1, p. 161-216. 2010.

CECI, Flávio. **UM MODELO SEMIAUTOMÁTICO PARA A CONSTRUÇÃO E MANUTENÇÃO DE ONTOLOGIAS A PARTIR DE BASES DE DOCUMENTOS NÃO ESTRUTURADOS**. 2010. 131 f. Dissertação (Mestrado) - Universidade Federal de Santa Catarina, Florianópolis, 2010

CHEN, Li; PU, Pearl. Critiquing-based recommenders: survey and emerging trends. **User Modeling And User-adapted Interaction**, S. I., v. 22, n. 2, p.125-150, abr. 2012.

CHEN, Rung-ching et al. A recommendation system based on domain ontology and SWRL for anti-diabetic drugs selection. **Expert Systems With Applications**, S. I., v. 39, n. 5, p.3995-4006, mar. 2012.

COLOMBO-MENDOZA, Luis Omar et al. **RecomMetz**: A context-aware knowledge-based mobile recommender system for movie showtimes. **Expert Systems With Applications**, S. I., v. 42, n. 3, p.1202-1222, 15 fev. 2015.

DAVIES, John; FENSEL, Dieter; VAN HARMELEN, Frank (Ed.). Introduction. In: DAVIES, John; FENSEL, Dieter; VAN HARMELEN, Frank. **TOWARDS THE SEMANTIC WEB: Ontology-driven Knowledge Management**. S. I.: Wiley, 2003. p. 1-10.

DEITEL, Paul J.; DEITEL, Harvey M. **Ajax, rich internet applications e desenvolvimento web para programadores**. São Paulo: Pearson Prentice Hall, 2008. xxiv, 747 p.

DEVEDZIC, Vladan et al. **Semantic Web and Education**. New York: Springer, 2006. 353 p.

DRACHSLER, Hendrik; HUMMEL, Hans; KOPER, Rob. Personal recommender systems for learners in lifelong learning: requirements, techniques and model. **International Journal Of Learning Technology**. S. I., p. 404-423. 12 maio 2007.

EGC. **Linhas de pesquisa do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento**. Disponível em <<http://www.egc.ufsc.br/index.php/pt/pesquisas/linhas-de-pesquisa>>. Acesso em 11 de fevereiro de 2014b.

FELFERNIG, Alexander et al. Developing Constraint-based Recommenders. In: RICCI, Francesco et al (Ed.). **Handbook on Ontologies**. New York: Springer, 2011. Cap. 6. p. 187-215.

FERDAOUS, Hdioud; BOUCHRA, Frikh; BRAHIM, Ouhbi. A comparison study of some algorithms in Recommender Systems. In: **INFORMATION SCIENCE AND TECHNOLOGY (CIST)**, 2., 2012, Fez. Proceedings... . S. I.: IEEE, 2012. p. 130 - 135.

FERNÁNDEZ, M. Overview of methodologies for building ontologies. In: **WORKSHOP ON ONTOLOGIES AND PROBLEM-SOLVING METHODS: lessons learned and future trends**. 1999. [S. l. : s. n.], 1999. p. 4-1, 4-13.

FERNÁNDEZ, M.; GÓMEZ-PÉREZ, A. Overview and analysis of methodologies for building ontologies. **Knowledge Engineering Rev.** v.17, n.2, 2002.

FREITAS, Frederico Luiz G. de. Ontologias e a Web Semântica. In: Renata Vieira; Fernando Osório. (Org.). **Anais do XXIII Congresso da Sociedade Brasileira de Computação**. Volume 8: Jornada de Mini-Cursos em Inteligência Artificial. Campinas: SBC, 2003, v. 8, p. 1-52.

FRIEDRICH, Gerhard; ZANKER, Markus. A Taxonomy for Generating Explanations in Recommender Systems. **Ai Magazine**, S. I., v. 3, n. 32, p.90-98, 2011.

GATZIOURA, Anna; SÀNCHEZ-MARRÈ, Miquel. A Case-Based Recommendation Approach for Market Basket Data. **Ieee Intelligent Systems**, v. 30, n. 1, p.20-27, fev. 2015.

GEDIKLI, Fatih; JANNACH, Dietmar; GE, Mouzhi. How should I explain? A comparison of different explanation types for recommender systems. **International Journal Of Human-computer Studies**, v. 72, n. 4, p.367-382, abr. 2014.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 5. ed. São Paulo: Atlas, 1999.

GOLDBERG, D. et al. Using collaborative filtering to weave an information tapestry. **Communications of the Association of Computing Machinery**, v. 35, n. 12, p. 61-70, 1992.

GÓMEZ-PÉREZ, Asunción. Ontological engineering: a state of the art. **Expert Update**, v. 2, n. 3, p. 33-43, 1999.

GÓMEZ-PÉREZ, A.; LÓPEZ, M. F.; CORCHO, O. **Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web**. Springer, 2002. 404 p.

GRANDISON, Tyrone. Security and Privacy in Web 2.0. **IEEE Internet Computing**, S. I., v. 18, n. 6, p.41-42, nov. 2014.

GROH, Georg; EHMIG, Christian. Recommendations in Taste Related Domains: Collaborative Filtering vs. Social Filtering. In: GROUP '07, 2007, Sanibel Island. **Proceedings of the 2007 International ACM Conference on Supporting Group Work**. New York: Acm, 2007. p. 127 - 136.

GRUBER, T.R. Towards principles for the design of ontologies used for knowledge sharing. **International Journal of Human-Computer Studies**, v. 43, p. 907-927, 1993.

GUARINO, N. Formal Ontology and Information Systems. In: **FORMAL ONTOLOGY IN INFORMATION SYSTEMS**, 1998, Trento. Anais... Amsterdam: IOS Press, 1998.

GUARINO, N., Oberle, D., Staab, S., What is an Ontology? In: **Handbook on Ontologies**, Second edition, Staab, S., Studer, R. (Eds.), Springer, 2009, pp. 1 – 17.

GUIZZARDI, G.; Almeida, João Paulo Andrade ; GUIZZARDI, R. S. S. ; FALBO, R. A.. Ontologias de Fundamentação e Modelagem Conceitual. In: **2o. Simpósio Brasileiro de Pesquisas em Ontologias (ONTOBRA 2009)**, 2009.

HAWALAH, Ahmad; FASLI, Maria. Utilizing contextual ontological user profiles for personalized recommendations. **Expert Systems With Applications**, S. I., v. 41, n. 10, p.4777-4797, fev. 2014.

HIRST, Graeme. Ontology and the Lexicon. In: STAAB, Steffen; STUDER, Rudi (Ed.). **Handbook on Ontologies**. 2. ed. Berlin: Springer, 2009. Cap. 12. p. 269-292.

FERDAOUS, Hdioud; BOUCHRA, Frikh; BRAHIM, Ouhbi. A comparison study of some algorithms in Recommender Systems. In: **INFORMATION SCIENCE AND TECHNOLOGY (CIST)**, 2., 2012, Fez. Proceedings... . S. I.: Ieee, 2012. p. 130 - 135.

FERNANDEZ-LOPEZ, M.; GOMEZ-PEREZ, A.; JURISTO, N. Methontology: From Ontological Art Towards Ontological Engineering. **Spring Symposium on Ontological Engineering of AAAI**. Stanford University, California, 1997. p. 33-40.

HERLOCKER J. L. et al. Evaluating collaborative filtering recommender systems. **ACM Transactions on Information Systems**, v. 22, n. 1, p. 5-53, 2004.

HUANG, A. Similarity Measures for Text Document Clustering. **Proceedings of the New Zealand Computer Science Research Student Conference 2008**. Christchurch , [S.n.] 2008.

HUANG, Shiu-li. Designing Utility-based Recommender Systems for e-Commerce: Evaluation of Preference-elicitation Methods. **Electronic Commerce Research And Applications**, Amsterdam, v. 4, n. 10, p.398-407, jul. 2011.

IKEDA, M., SETA, K., KAKUSHO, O., et al., Task ontology: Ontology for building conceptual problem solving models, In: **Proceedings of ECAI98 Workshop on Applications of ontologies and problem-solving model**, pp.126-133, 1998.

JESUS, Deise Lourenço de; CUNHA, Murilo Bastos da. Produtos e serviços da web 2.0 no setor de referência das bibliotecas. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 17, n. 1, p. 110-133, jan./abr. 2012.

JONES, M. Tim. **Recommender systems, Part 1**: Introduction to approaches and algorithms. 2013. Disponível em: <<http://www.ibm.com/developerworks/library/os-recommender1/>>. Acesso em: 12 maio 2014.

KAGIE, Martijn; VAN WEZEL, Michiel; GROENEN, Patrick J.f.. Map Based Visualization of Product Catalogs. In: RICCI, Francesco et al. **Recommender Systems Handbook**. New York: Springer, 2011. Cap. 17. p. 546-576.

KIM, Byeong Man et al. A new approach for combining content-based and collaborative filters. **Journal Of Intelligent Information Systems**, S. I., v. 27, n. 1, p.79-91, jul. 2006.

LOHR, Steve. A \$1 Million Research Bargain for Netflix, and Maybe a Model for Others. **The New York Times**. New York. 21 set. 2009. Disponível em: <<http://www.nytimes.com/2009/09/22/technology/internet/22netflix.html?pagewanted=all>>. Acesso em: 18 jun. 2014.

MARTINS, Aline Freitas ; FALBO, R. A. . Models for Representing Task Ontologies. In: **3rd Workshop on Ontologies and their Applications - WONTO'2008**, 2008, Salvador. Proceedings of the 3rd Workshop on Ontologies and their Applications, 2008.

MATTOS, Merisandra Côrtes de; SIMÕES, Priscyla Waleska Targino de Azevedo; FARIAS, Renan Figueredo. A METODOLOGIA METHONTOLOGY NA CONSTRUÇÃO DE ONTOLOGIAS. **Revista de Iniciação Científica**, Criciúma, v. 8, n. 1, p.1-12, jan. 2007.

MELVILLE, Prem; SINDHWANI, Vikas. Recommender Systems. In: SAMMUT, Claude; WEBB, Geoffrey I.. **Encyclopedia of Machine Learning**. S. I.: Springer, 2011. p. 829-838.

METTOURIS, Christos; PAPADOPOULOS, George A.. Ubiquitous recommender systems. **Journal Computing**, Vienna, v. 96, n. 3, p.223-257, 1 mar. 2014.

MIDDLETON, Stuart E.; ROURE, David de; SHADBOLT, Nigel R.. Ontology-Based Recommender Systems. In: STAAB, Steffen; STUDER, Rudi (Ed.). **Handbook on Ontologies**. 2. ed. New York: Springer, 2009. p. 779-797.

MIZOGUCHI, R.; VANWCKELKENHUYSEN, J., IKEDA, M. Task ontology for reuse of problem solving knowledge. In: **ECAI'94 TOWARDS VERY LARGE KNOWLEDGE BASES**. Proceedings... Amsterdam: IOS Press, 1995. p. 46-59.

MORID, Mohammad Amin; SHAJARI, Mehdi; HASHEMI, Ali Reza. Defending recommender systems by influence analysis. **Information Retrieval**, v. 17, n. 2, p.137-152, 1 abr. 2014.

NOY, N. F.; MCGUINNESS, D. L. **Ontology Development 101: A Guide to Creating Your First Ontology**. Stanford University, p. 1-25, 2001.

O'REILLY, Tim. **What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software**. 2005. Disponível em: <<http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>>. Acesso em: 12 set. 2014.

OLIVEIRA, Viviane Nogueira Pinto de; ALMEIDA, Mauricio Barcellos. Um roteiro para avaliação ontológica de modelos de sistemas de informação. **Perspect. ciênc. inf.**, Belo Horizonte , v. 16, n. 1, mar. 2011 .

OLIVEIRA, Hellen Carmo de; CARVALHO, Cedric Luiz de. **Gestão e Representação do Conhecimento**. Goiás, 2008. 20 p. Disponível em: <http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_003-08.pdf>. Acesso em: 20 nov. 2014.

OLIVEIRA, Lucas A. de; FIGUEIRA, Rui A. R. B.; LOPES, Expedito Carlos. Aplicação de um Metamodelo de Contexto a uma Tarefa de Investigação Policial. In: SEMINAR ON ONTOLOGY RESEARCH IN BRAZIL AND INTERNATIONAL WORKSHOP ON METAMODELS, ONTOLOGIES AND SEMANTIC TECHNOLOGIES, 5., 2012, Recife. **Proceedings...** . Anais, 2012. p. 284 - 289.

PAIVA, Fábio Augusto Procópio de; COSTA, José Alfredo Ferreira; SILVA, Cláudio Rodrigues Muniz. A Hierarchical Architecture for Ontology-based Recommender Systems. In: **BRAZILIAN CONGRESS ON COMPUTATIONAL INTELLIGENCE**, 11., 2013, Porto de Galinhas. **Proceedings...** . S. I.: Anais, 2013. p. 362 - 367.

PICAULT, Jérôme et al. How to Get the Recommender Out of the Lab? In: RICCI, Francesco et al. **Recommender Systems Handbook**. New York: Springer, 2011. Cap. 10. p. 324-265.

Pouchard, L., Ivezic, N., and Schleno, C. (2000). Ontology engineering for distributed collaboration in manufacturing. In: **Proceedings of the Association for Information Systems (AIS) Conference**. Citeseer.

RAUTENBERG, Sandro et al. Uma metodologia para o desenvolvimento de ontologias. **Revista Ciências Exatas e Naturais**, A, v. 10, n. 2, p.237-262, 2008.

RESNICK, P.; IACOVOU, N.; SUCHAK, M.; BERGSTORM, P.; RIEDL, J. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In: **Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work**, 1994, Chapel Hill, North Carolina. Anais. ACM. p.175–186.

RICCI, Francesco et al (Ed.). **Recommender Systems Handbook**. New York: Springer, 2011. 842 p.

RICCI, Francesco; ROKACH, Lior; SHAPIRA, Bracha. Introduction to Recommender Systems Handbook. In: RICCI, Francesco et al (Ed.). **Recommender Systems Handbook**. S. I.: Springer, 2011. p. 1-35.

SAAD, Beth. **Estratégias 2.0 para a mídia digital**: internet, informação e comunicação. 2. ed. rev. e atual. São Paulo (SP): SENAC, 2008. 218 p.

SCHAFER J. B; KONSTAN J. A; RIEDL J. E-Commerce Recommendation Applications. **Data Mining and Knowledge Discovery**. v. 5, n 2, p. 115-153, 2001.

SCHRODER, G.; THIELE, M.; LEHNER, W. Setting Goals and Choosing Metrics for Recommender System Evaluations. **UCERSTI Workshop at the ACM Conference on Recommender Systems**, 2011.

SANTAROSA, Lucila Maria Costi; CONFORTO, Débora; BASSO, Lourenç O de Oliveira. Ferramentas de autoria e de colaboração: discutindo a acessibilidade e a usabilidade na perspectiva da Web 2.0. **Revista Brasileira de Informática na Educação**, Porto Alegre, v. 21, n. 1, p.122-132, 25 jun. 2013.

SCHULZ, Stefan et al. Vantagens e limitações das ontologias formais na área biomédica. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, Rio de Janeiro, v. 3, n. 1, p.33-48, 3 abr. 2009

SENECAL S. NANTEL J; The influence of online product recommendations on consumers' online choices. **Journal of Retailing**, v. 80, n. 2, p. 159-169, 2004.

SEGARAN, Toby. **Programando a inteligência coletiva**: desenvolvendo aplicativos Web 2.0 inteligentes. Rio de Janeiro : Alta Books, 2008.

SEWALD JUNIOR, E. ; ROTTA, M. J. R. ; SILVA, E. R. G. ; ROVER, A. J. ; GAUTHIER, F. A. O. ; CARMO, V. M. . Processo de Desenvolvimento de Ontologia de Domínio da Área Jurídica: Caso Acidentes de Trânsito. In: **8ª Conferência Ibérica de Sistemas e Tecnologias de Informação**, 2013, Lisboa. Atas da 8ª Conferência Ibérica de Sistemas e Tecnologias de Informação. Lisboa: Associação

Ibérica de Sistemas e Tecnologias de Informação, 2013. v. II. p. 212-217.

SHARDANAND, U.; MAES, P.(1995) Social Information Filtering: Algorithms for Automating “Word of Mouth”. In: **ACM CHI’95 CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS**. Proceedings... . [S.l.: s.n.], 1995. v.1, p.210–217.

SILVA, E. L. da; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. 3. ed. rev. atual. Florianópolis: Laboratório de Ensino a Distância da UFSC, 2001.

SILVA, Daniela Lucas da; SOUZA, Renato Rocha; ALMEIDA, Maurício Barcellos. Ontologias e vocabulários controlados: comparação de metodologias para construção. **Ciência da Informação**, Brasília , v. 37, n. 3, p. 60-75, dez. 2008 .

STAAB, Steffen; STUDER, Rudi. **Handbook on ontologies**. 2nd ed. Berlin; Springer, 2009. xix, 811 p.

STUDER, Rudi; BENJAMINS, V. Richard; FENSEL, Dieter. Knowledge Engineering: Principles and methods. **Elsevier Data & Knowledge Engineering**, 1998. p. 161-197.

TINTAREV, Nava; MASTHOFF, Judith. Designing and Evaluating Explanations for Recommender Systems. In: RICCI, Francesco et al. **Recommender Systems Handbook**. New York: Springer, 2011. p. 479-510.

TORRES, Roberto. **Personalização na Internet**: como descobrir os hábitos de consumo de seus clientes, fidelizá-los e aumentar o lucro de seu negócio. São Paulo: Novatec, 2004.

VALOIS B. JR, Cleomar; OLIVEIRA, Marcius Armada de. Recommender systems in social networks. **Journal of Information Systems and Technology Management**, São Paulo, v. 8, n. 3, Dec. 2011

VESIN, Boban et al. Ontology-Based Architecture with Recommendation Strategy in Java Tutoring System. **Computer Science and Information Systems**, S. I., v. 10, n. 1, p. 237-261, jan. 2013.

W3C. **OWL 2 Web Ontology Language**: Structural Specification and Functional-Style Syntax. 2009. Disponível em:
<<http://www.w3.org/2007/OWL/wiki/Syntax>>. Acesso em: 10 jan. 2015.

WALK, Simon et al. Recommending Concepts to Experts: An Exploration of Recommender Techniques for Collaborative Ontology Engineering Platforms in the Biomedical Domain. **Ceur Workshop Proceedings**. [S. L.], 2012.

WALTER, Frank E. et al. Moving recommender systems from on-line commerce to retail stores. **Information Systems And E-business Management**, S. I., v. 10, n. 3, p.367-393, set. 2012.

WANG, X., CHAN, C.W., Ontology Modeling in UML. In **Proceedings of International Conference on Object Oriented Information Systems 2001**, Calgary, Canada, Springer-Verlag, 59-68, 2001.